Powerful Priorities or Inbuilt Incentives?

The Impact of Competitive Research Funding on the Supply of Science

Emil Bargmann Madsen

Powerful Priorities or Inbuilt Incentives?

The Impact of Competitive Research Funding on the Supply of Science

Ph.D. dissertation

Politica

ISBN: 978-87-7335-283-0

Cover: Svend Siune
Print: Fællestrykkeriet, Aarhus University
Layout: Annette Bruun Andersen

Submitted July 12, 2021
The public defence takes place December 8, 2021
Published December 2021

Forlaget Politica
c/o Department of Political Science
Aarhus BSS, Aarhus University
Bartholins Allé 7
DK-8000 Aarhus C
Denmark

"Without doubt, the most abnormal thing in this age of Big Science is money. The finances of science seem highly irregular
and since they dominate most of the social and political implications, our analysis must start here."

- Derek J. de Solla Price
Little Science, Big Science... and Beyond (1986, p. 82)

Acknowledgements

THIS thesis was initially about cumulative advantage, or the notion that once you are rich you tend to become richer more easily. This is also true for writing a dissertation. Once someone offers their advice and help, you tend to quickly accumulate more and end up with an enormous debt of gratitude towards the people who have selflessly helped you become richer in experience. Large parts of this debt is due to my advisors Kaare Aagaard, Carsten Jensen, and Jens Peter Andersen. Kaare has tirelessly listened to a host of my terrible and less-terrible ideas, and often responded with the succinct suggestion of starting over with a blank page. While initially frustrating, I think this dissertation would have turned out much worse, or not at all, had he not insisted on this and many other great pieces of advice. The fact that he has retained faith in the overall project is a testament to his abilities as a supervisor, and an invaluable motivation. Carsten has provided me with guidance through many years, and I am grateful that he agreed to continue doing so even if I have strayed far of his area of expertise. His close attention to what works and what does not, and his ability to think about process, not only outcome, has been an immense source of support. Jens Peter has probably helped me in more ways than I can remember. Everything from navigating the world of scientometrics to providing me with increasingly more data, Jens Peter always set aside time to discuss my weird plans for some obscure analysis.

In 2019, I was fortunate enough to spend time at the Science Policy Research Unit (SPRU) at the University of Sussex. Michael Hopkins graciously agreed to host me, provide excellent feedback, and introduce me to the ambitious and exciting research environment that is SPRU. I am also greatful for the many great talks and conversations I had with Frédérique Bone, Daniele Rotolo, and Ohid Yaqub. I am further indebted to many great Ph.D.-students at SPRU, who I had the pleasure of sharing an office with during my stay.

While time spent abroad is valuable to any Ph.D.-project, I would be amiss if I did not acknowledge all the help provided to me by colleagues and fellow Ph.D.-students from the Danish Centre for Studies in Research and Research Policy (CFA) and the Department of Political Science in Aarhus. My office mates, Thomas, Alexander, Christine, and Simon have each contributed just the right amount of humour, laughter, engaging conversation, and coffee breaks I needed to finish this dissertation. Thomas, Andreas, Alexander and Ea all deserve the highest recognition for their encouragement, guidance, and general willingness to help in any way possible. With friends like you, this dissertation seems like only a minor outcome of the past four years. The same goes to the rest of the small, but growing, pack of Ph.D.-students and research assistants at CFA. Thank you to Allan, Peter, Massimo, Asger, Anna-Kathrine, Emer, Astrid, Josefine, and all the rest. I am also greatly indebted to especially Jesper Wiborg Schneider, Carter Bloch, Philippe Mongeon, Lise Degn, and Mathias Wullum Nielsen for ever more engaging conversations and advice, which helped move the project forward, and give me hope for my career after this. However, no one should receive higher praise than our centre secretary Jane who, after all, is the colleague none of us could do without - me especially. At the Department of Political Science, I have met an incredible bunch of young researchers, who are too numerous to be mentioned in total. But I want to acknowledge at least Andreas, Lauritz, Sarah, Mathilde, Frederik, Suthan, Niels, Dani, Matias, Jonas, Rasmus, Alexander, and Mathias. Thank you!

Lastly, I am particularly grateful for the support and encouragement from my family. Above all others, Lene has been the most supportive friend and partner one could even hope for. I am not quite sure how she managed to stay positive, encouraging, and (extremely) tolerant during the past four years. I am sure that I would never have finished this dissertation, had it not been for all her sacrifices.

Emil Bargmann Madsen
Silkeborg, July 2021

Preface

THIS report forms part of my Ph.D. project ”Powerful Priorities or Inbuilt Incentives? The Impact of Competitive Research Funding on the Supply of Science” conducted at the Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University. Within this project, I study how competitive research funding is distributed across both individual researchers, the topics and disciplines they investigate, and the possible drivers of these allocative patterns. The project consists of this report along with four self-contained research articles:

A. Madsen, Emil Bargmann. Diversity or Disparity? The Concentration of Funded Research Topics in the United Kingdom. *Under review.*

B. Madsen, Emil Bargmann and Kaare Aagaard (2020). Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers, *Quantitative Science Studies*, 1(3), 1159-1189.

C. Madsen, Emil Bargmann and Mathias Wullum Nielsen. Thematic funding and topic switching in a sample of 10,475 scientists. *Under review.*

D. Madsen, Emil Bargmann. The Interdependence of Research Funding Concentration, Policy Priorities, and Problem Choice. *Working paper.*

The following report presents a general discussion of the drivers of funding concentration in relation to research topics and disciplines, and a synthesis model relating the contributing factors outside and inside the scientific system to each other. I draw on theoretical ideas and empirical results from all of the four articles, but the argument cuts across these and provides a framework for understanding research funding concentration that goes beyond any single paper. The purpose of this report is further to synthesise the empirical results from each of the articles and discuss these against prominent debates within science and research policy. Questions on how to design research funding instruments and balance funder priorities often lack solid empirical footing. I hope to contribute with at least some basis for coming policy discussions and point to important areas with a shortage of even basic knowledge on how and why funding is concentrated.

Contents

1. Introduction

One of the most important research policy developments since World War Two has been the transition from exclusive block funding of research organisations and universities to a split funding mode, where yearly lump sum funding is combined with competitive project funding (Gläser & Laudel, 2016, p. 122). The split mode funding of research started out with the setup of research councils as quasi-public organisations tasked with allocating a portion of the research budget, and was quickly adopted or captured by scientists as legitimate instruments for distributing resources in the scientific system (Rip, 1994). Other funding organisations have since joined the traditional research councils as sources of external project funding, including a host of governmental funding bodies, charities, and private research foundations (Gläser & Velarde, 2018; Sörlin, 2007). Together, funding deriving from external sources comprises a much larger relative share of total funding now than ten or twenty years ago (Whitley et al., 2018). In Denmark, the share of competitive funding has grown from around 10 percent to almost 50 percent of the national research budget (Aagaard, 2017; Lind et al., 2019; The Danish Council for Research and Innovation Policy, 2020a; The Danish Council for Research and Innovation Policy, 2020b). A larger share of public funding is also being awarded through open competition, and private and charitable research foundations are awarding a much larger amount of funding than hitherto. External sources of research funding have accordingly become integral to the conduct of modern science.

Competitive research funding has had the dual purpose of allocating scarce resources based on a meritocratic model of peer-review, and introducing an element of competition to ensure that the best ideas and projects are rewarded with resources to carry out the work. Competitive funding schemes have thus become both internal quality assurance and reward systems in themselves (Aagaard, 2017, pp. 75–76), and have been promoted in the hope that funding allocations guided by the scientific community itself will accurately identify the most promising topics and scientists. The hope is that allocation of funds guided by the scientific opinion of peers will effectively weed out ideas not worth pursuing, and "the distribution of grants will automatically yield the maximum advantage for the advancement of science as a whole" (Polanyi, 1962, p. 62).

However, as science has outgrown competitive funding budgets, research councils and government funders have also begun incorporating specific policy priorities in evaluation and funding procedures. Increased selectivity and targeting of research funding towards particular research areas or problems have added to the growing competition for funding between both researchers and scientific fields (Gläser & Laudel, 2016; Kearnes & Wienroth, 2011).

Originally, peer review of funding applications had merely been an optional addendum to the allocation process, but growing demands for accountability have led scientists to recast peer-review as a necessary part of grant application. Competitive funding schemes, guided by evaluation from experts, are viewed as crucial for selecting the best proposals and ensuring the credibility of science (Baldwin, 2018). However, as the system for scientific funding have grown and evolved, many now fear that it has brought along some unintended consequences directly counterproductive to these goals. The introduction of competitive research funding has added multiple layers to the scientific system and changed its dynamic. Researchers across many fields have become reliant on acquiring funds from external sources and via competitive funding schemes. Some are entirely dependent on winning grants, not only to carry out work within their chosen topic, but to continue working altogether (Gläser & Laudel, 2016, p. 122; Whitley et al., 2018, p. 110).

One of the most visible and debated consequences of greater reliance on competitive research funding has been an increasing degree of funding concentration[1]. A growing body of both science commentary and empirical studies have highlighted the tendency for competitive research funds to be concentrated in the hands of a few individual researchers or research groups (Aagaard et al., 2020; Larivière et al., 2010; Mongeon et al., 2016; Wahls, 2016; Wahls, 2019). Studies of funding concentration in Denmark (Aagaard et al., 2019, p. 33), the United States (Katz & Matter, 2020; Wahls, 2018a), the United Kingdom (Ma et al., 2015), Quebec (Larivière et al., 2010; Mongeon et al., 2016), and China (Wu, 2015; Zhi & Meng, 2016) show very similar patterns. Around 10 % of researchers and research institutions are awarded 40 % of the total amount of funding, while 20-25 % control 70-80 %[2].

[1]As a side note, "funding concentration" refers to the trend of allocating larger shares of funding or number of grants to fewer individuals, groups, and units, while "funding dispersal" reference the distribution of smaller shares (adapted from Aagaard et al., 2020, p. 1).

[2]The distributions differ between countries and disciplines. In Quebec, 11-18 % of researchers are principal investigators on around 80 % of external grants depending on their discipline (Larivière et al., 2010; Mongeon et al., 2016). In the biomedical field, only 55 % of researchers had at least one grant during the 2000-2007 period, while 75 % of social scientists had access to external grants (Larivière et al., 2010, p. 48). Similar patterns have been observed across 12 public and private research funders in Denmark. In the Natural, Technical, and Medical Sciences, 20 % of grantees are awarded 55 % of all grants whereas they command only 40 % in the Social Sciences and Humanities (Aagaard et al., 2019, p. 33).

Some studies have even documented a pronounced increase or persistence in concentration patterns across time and within different national and disciplinary settings (Katz & Matter, 2020; J. Li et al., 2017; Ma et al., 2015). The large, and sometimes increasing, degree of concentration in research funding is likely rooted in a series of changes to contemporary external funding systems. Research funders in both the public and private sector appear to have moved towards awarding a smaller number of grants, while simultaneously increasing the monetary value or size of individual awards (Bloch & Sørensen, 2015; Hendricks, 2019; Ma et al., 2015). Many public funders, across countries, have introduced "Excellence"-oriented grants that focus on funding top performing researchers, or establishing whole research centres focused on a small segment of high performing research groups (Aagaard, 2017; Bloch et al., 2016; Bloch & Sørensen, 2015; Ida & Fukuzawa, 2013).

Stronger concentration of competitive research funding is most likely also a function of the tendency for the same scientists to win multiple grants, and not just larger grants. The trend is not widely documented as a general pattern, but evidence from some research funding bodies suggests that this tendency is prominent (Larivière et al., 2010; Mongeon et al., 2016). For example, some of the most well-funded applicants to the U.S National Institutes of Health (NIH) simultaneously manage more grants than the average applicant would win in their entire career, with some top-funded researchers having more than eight running NIH grants (Hand, 2012; Hand & Wadman, 2008).

The existing evidence mainly focus on concentration on individual researchers, but an underlying theme or assumption is that strong concentration may exclude valuable perspectives on what is being studied (e.g. Alberts et al., 2014). Most empirical research have treated funding concentration at the individual and research group level as a potential problem for the efficiency of scientific discovery. Strong concentration of funding seems to lead to diminishing marginal returns, where cumulative funding amounts above a certain threshold do not necessarily translate into more publications, citations, or higher impact (Berg, 2013; Bonaccorsi & Daraio, 2005; Fedderke & Goldschmidt, 2015; Fortin & Currie, 2013; Gök et al., 2015; Hernandez-Villafuerte et al., 2017; M. Lauer et al., 2017; M. S. Lauer et al., 2015; Lorsch, 2015; Mongeon et al., 2016; Nag et al., 2013; Wadman, 2010; Wahls, 2018b; Wahls, 2018a). Concentration may further inhibit the creation of a scientific growth layer where future top scientists can be recruited from (Aagaard et al., 2020; Bloch & Sørensen, 2015).

One can worry about how different degrees of concentration or dispersal impact the amount of publications or citations produced by researchers in different strata of the resource distribution. But perhaps there are other equally important research funding policy perspectives in relation to concentration, which so far have received too little attention? Discussions of competitive research funding's ability to foster scientific advancement and even breakthroughs have so far been overly focused on how grant funding is transformed into publications and citations. Less focus (however see Azoulay et al., 2019; D. Li et al., 2017) has been granted to the question of how competitively funded research is translated into relevant knowledge, tangible products and practical solutions for global grand challenges. This entails discussing how research funding policies can help ensuring the right balance of research across different problem areas. In order to qualify this discussion, we need to examine what type of research is actually being funded. How funding policy impacts what individual researchers choose to pursue, and how it adds up to system level shifts or stand-stills is not yet well-known (Gläser & Laudel, 2016, p. 133). The fundamental goal of this dissertation is to contribute to this objective: A main aim is accordingly to shift focus from who receives funding (i.e. persons, groups, institutions) to what is funded.

However, unlike studying who receives what amount of competitive research funding, the examination of the aboutness or the content of what is funded is a much less straightforward unit of analysis. Whereas individual researchers are self-contained units who are members of certain research groups, and employed by specific research organisations, research content is far less tangible. Research content instead refers to more loose concepts such as research fields, disciplines, specialties, and topics. These are vague constructs describing different pieces of research that go together, or whose aboutness is similar in some way. Here, I make use of the term 'research content' to describe aggregate constructs of what differing grants or funded research projects are about, and terms such as "fields", "disciplines", "specialties", and "topics" as categories of research content hierarchically ordered from broad to narrow.

1.1 Research question and approach

A key overall question in the thesis is how competitive research funding is guiding research content? Drawing on analyses of the scholarly literature and few funding organisations, previous studies have identified two seemingly consistent trends. First, what is being studied is often rather narrow and static. Studies of the biomedical sciences find that the topics and specialties funded the most, are already well-established in the literature, and have received a lot of funding in the past (e.g. Edwards et al., 2011; Klavans & Boyack, 2017a; Stoeger et al., 2018; Stoeger & Nunes Amaral, 2020). Second, studies of disease-specific research suggests that the type of research published and funded often do not correspond with the most pressing problems, or follow potentially optimal strategies[3] for discovery and breakthroughs (Foster et al., 2015; Gerow et al., 2018; Yao

[3] In this instance, the "optimal" strategy refers to an optimized search strategy when discovering the network of chemical entities published in journal articles (Rzhetsky et al., 2015).

et al., 2015). Also, there tends to be a mismatch between diseases that are highly investigated and funded, and diseases with a high societal health burden (Evans et al., 2014; Ràfols & Yegros-Yegros, 2018; Yegros-Yegros et al., 2018; Yegros-Yegros et al., 2020).

This path dependency of funding distributions in the biomedical field may be intimately linked to concentration of funding on individual researchers. At least, tentative findings show that topic choice is one predictor of funding success in the U.S. National Institutes of Health (NIH) (Hoppe et al., 2019; Stoeger et al., 2018). Funding opportunities are thus often mentioned as a key reason for a path dependency of the type of research conducted, but it has seldom been empirically tested outside the biomedical field.

To foster a discussion on how competitive research funding shapes the conduct of research beyond differences in publication output and citations, a first necessary step is to improve evidence on how funding is distributed across different types of research content. The aim of this dissertation is therefore to shed light on how competitive funding influences the direction of research by examining:

> *1. How is competitive research funding distributed across research content at different levels of granularity?*

Despite increased attention to the question of what type of research content is funded outside of the biomedical sciences, we still know very little about the extent of funding concentration on different types of research content such as research disciplines and specialties or the more narrow topics scientists engage in. Answering this question is an important basis for discussing priorities, and how to balance funding allocations towards diverse research problems. Funding decisions are made by multiple separate research funding organisations, both publicly and privately financed, where each funding decision is only compared to decisions within the same call for applications. Furthermore, individual grant reviewers may only have knowledge of the projects they helped evaluate. Assessing what research is funded at an aggregate level can help illuminate possible imbalances or patterns not realised when funding decisions are decentral and without a clear portfolio overview. This ties closely to past and current discussions of how research funding should "ideally" be allocated (Bozeman, 2020; Sarewitz & Pielke, 2007; Yao et al., 2015). Arguably, some areas of research are more important than others. Accordingly, a skewed focus on a smaller segment of research problem is natural (Edwards et al., 2011). Some inequality in the distribution of attention and resources is thus warranted and expected given differences in importance. However, too strong concentration is often criticised for its assumed effects on the conduct of science (Aagaard et al., 2020). Strong concentration will generally leave some science undone (Aagaard et al., 2020, pp. 11–12; Gläser & Laudel, 2016, p. 152), and may deter science from addressing substantial demands from society (Ciarli & Ràfols, 2019; Evans et al., 2014; Yao et al., 2015).

Further, exploring what type of research is being funded may also inform discussions of how current patterns of funding concentration or dispersal arise and may provide insights into why some research topics are prominent while others appear relatively underfunded (C. Gross et al., 1999; Hanna, 2015; Waltman et al., 2019). I therefore supplement the more descriptive question above, with a more explanatory one:

> *2. What are the potential drivers of research funding concentration or dispersal across research content?*

The expectation is that a broad-based analysis of what type of research is funded across national, disciplinary, and funder contexts can help illuminate potential drivers behind allocative patterns. For example, is strong funding concentration on individual researchers related to concentration at the content level? Have governments' use of targeted funding mechanisms served to create more or less diversity in what problems researchers address? What is the role of privately funded grants in shaping the direction of research undertaken by researchers in publicly funded universities?

Understanding funding patterns and exploring its potential causes brings us closer to understanding the consequences of how competitive research funding is institutionalised and how policy decisions can affect the direction of research. This more explanatory question is however highly complex and cannot be definitively answered within the scope of this thesis. However, some preliminary steps will be taken towards identifying and disentangling key drivers and their potential interactions, which may result in both positive and negative reinforcement depending on the circumstances.

1.1.1 Answering the research questions

Despite the previous overt focus in the scientific literature on how funding is distributed across individual researchers, this thesis does not represent the first attempt to understand what type of research is funded. However, existing work still leaves many questions unanswered. Extant investigations largely analyse singular funding organisations, such as the NIH, and mostly in a North American context (Klavans & Boyack, 2017a; Manton et al., 2009; Stoeger et al., 2018; Yao et al., 2015). Not much focus has yet been placed on the possible similarities or differences across national contexts or how the interplay between funding organisations, both private and public, affect the distribution of funding.

To study the ramifications of increased funding concentration on what research is studied, I combine existing and new data on individual grants awarded to researchers by

22 public and private research councils and research foundations in Denmark and the United Kingdom, from 2005-2017, with research publications published by the principle investigators of these grants.

The main contribution of the dissertation is the descriptive evidence of funding allocations across different types of research content, and in highlighting what type of research is highly or less highly funded. Categorising different research grants according to what they are about is a first prerequisite for describing these patterns.

Previous research has made several attempts at describing what type of research is being funded by using the textual descriptions grant recipients include in their applications, such as the project titles or project summaries. Some have provided incredibly detailed accounts by partitioning 364,000 U.S. federally funded grants into more than 90,000 research topics based on their textual similarity to research articles (Klavans & Boyack, 2017a, p. 1166). Similar levels of detail have been achieved by estimating funding allocated to research into different genes (Stoeger et al., 2018). Here, categorisation of funding rested on what genes were addressed in the publications a grant led to (Stoeger et al., 2018, p. 15). Other studies have resorted to broader categories of research content, such as the medical specialities addressed in funding applications to the NIH, by submitting application texts to natural language processing (Hoppe et al., 2019). Finally, the distribution of funding to different disciplines in Quebec, has been assessed through categorising grants by the departmental names of their applicants (Larivière et al., 2010).

To understand how research funding in Denmark and the United Kingdom is distributed, I go beyond these studies in two distinct ways. Firstly, I make an attempt at categorising research grants at multiple levels of content aggregation, both the more granular topical or speciality level and at the broader disciplinary and field level. In chapter 2, I lay out how these ”levels of analysis” differ. Secondly, I use research articles published by grant recipients as a way to categorise individual grants on the basis of the output it entails, rather than the intention as stated in application abstracts and titles. These categorisations of the content in research grants form the basis of a wide-ranging descriptive analysis of what type of research is funded in Denmark and the UK, when taking into account a wide range of research funding organisations.

I show that funding is generally skewed across different categories of research content. Funding provided by both public research councils and private research foundation tends to cluster within a small set of disciplines, a select group of research topics, and within a narrow set of disease areas. This strong degree of concentration provides little basis for explanation in itself. It may reflect that researchers and funders focus their efforts and resources on the most promising and worthwhile topics, and that science functions as self-organising system, where the individual rational pursuit of important research problems guides the collective behaviour of scientist (Polanyi, 1962). However, it may also reflect external policy priorities.

I use these descriptive findings as an outset for exploring and discussing some possible drivers of concentration. Relatively little is known about how the organisation of competitive research funding shapes the direction and content of research (Braun, 1998; Gläser & Laudel, 2016, p. 125). The literature has tended to shy away from overtly explaining why funding may concentrate in selected topics and disciplines, beyond the notion that researchers tend to react to what their peers perceive as important and where opportunities for new discoveries exist. However, the existing evidence points towards especially two type of explanatory factors.

The first focuses on the *internal* norms and institutions for allocating rewards and recognition in the scientific system. Heavily funded topics are highly visible and prestigious, and may attract relative more funding because of pre-existing prominence within the scholarly literature and high levels of funding in the past (Edwards et al., 2011; Klavans & Boyack, 2017a; Stoeger et al., 2018). Funders use peer-review to discriminate between funding applications, and these decision-making processes have been argued to impact the content of funded research. Reviewers may be prone to selectively define what and who can be considered excellent and "fundable". Also, researchers themselves may adapt what they study to accommodate reviews (Gläser & Laudel, 2016, pp. 123–124).

To provide an empirical basis for these discussions, I analyse whether well-funded topics and specialities exhibit signs of positive feedback, where high or low levels of funding in the past tends to result in high or low levels of funding in subsequent years. I also prod the proposed relationship between concentration on individual researchers and concentration in research content. When a small section of funded researchers also receives the bulk of funding amounts, what happens to the type of research funded? Using funding data from both Denmark and United Kingdom my results suggests that researchers tend to stay within a narrow selection of topics, and that well-funded researchers are more likely to research already well-funded topics, specialities, and disciplines.

The second type of explanation instead recognises that the priorities of scientists are not isolated from the *external* priorities of the funders providing the resources. Governments and public research funders are increasingly using earmarked funding to direct the content of research to areas of particular interest (Hegde & Mowery, 2008; Hegde & Sampat, 2015; Myers, 2020). Similarly, private funders are playing an increasingly prominent role in funding publicly employed scientists, and can be selective in the type of research they sponsor. Increased reliance on these funding sources may shift the incentive to study a narrower or broader set of topics or disciplines. Firstly, I compare allocative patterns from funding provided through responsive mode grants, where researchers are free to choose their problem of study, versus more strategic or targeted grants, which include some fixed problem areas, or where studies on certain topics are encouraged. At the level of the funding systems in Denmark and the United Kingdom, a large degree of overlap exists between priorities in responsive mode and targeted mode grants. At the individual level, researchers receiving targeted grants are more prone to change their area of study in response, but often return to a set of topics they were already engaged in researching. Secondly, I ask whether private sector priorities act as a defining driver of the allocation of public grants, even beyond what they themselves allocate as argued in studies of the scholarly literature (Evans et al., 2014; Yegros-Yegros et al., 2020). Comparing publicly and privately funded grants also show a large overlap, and areas with distinct interest to Danish private research foundations are also highly funded through public means. I show that grants awarded by private funders, allocating funds within a narrow range of topic areas, seem to spill-over and guide what is funded through responsive mode grants. Priorities tend to overlap, because successful applicants draw funding from multiple sources, and highly funded individuals benefit from both private and public sources of competitive research funding.

In summary, this dissertation is an attempt to touch upon how competitive research funding serves to shape the direction of research conducted by focusing on both the descriptive patterns and trends, and using a rich material on individual research grants to propose some relevant drivers contributing to these patterns.

The conclusions of this dissertation draw on data on more than 120,000 individual research grants awarded by public and private funders in Denmark and the United Kingdom, and the scholarly output produced by each grantee. Each country has received less attention in previous studies of funding concentration, and also offers distinct opportunities for empirical analysis. Denmark has a very diverse system of funders with many private and public funding organisations, which makes it a possible setting for empirically analysing how public and private funding priorities interact and collectively impact the funding distributions. The United Kingdom, on the other hand, provides a suitable test bed for studies of positive and negative feedback in funding and examinations of what individual researchers choose to study within a funded project. The public research councils diligently track the scholarly output from awarded grants, which makes it possible to determine the content of scholarly output produced by individual researchers before, during, and after obtaining a competitive grant. Combining a study of funding from Danish and British funding organisations also contribute to illuminate how drivers of funding concentration may differ across different contexts. The purpose is however not to conduct a full comparative study of similarities and differences in funding concentration and its drivers in the two countries. Instead, the two county contexts offer different opportunities in illuminating different aspects of the overall research questions.

1.1.2 Structure of the dissertation

In this chapter, I have argued that besides examining concentration of competitive research funding at the individual level, there are good reasons to also exploring how funding is spread across different types of research. From this point of departure, this first chapter has laid out the main research questions guiding the dissertation and briefly outlined how the empirical contributions of the dissertation adds to providing some tentative answers.

In chapter 2, I provide a theoretical discussion, definition, and conceptualisation of research content. I discuss what is meant by the research content or "aboutness" of a research grant, and suggests some theoretical guidance for how to capture this content in empirical data. I also highlight some of the important methodological challenges and shortcomings related to this exercise. In chapter 3, I argue that we can think of research funding distributions as a question of supply and demand for different types of science. I then proceed to discuss three different frameworks for understanding how supply and demand interact within competitive funding systems, related to the possible drivers outlined above, and present a synthesis model relating these to the incentives for studying certain research content. Chapter 4 proceeds to present an overview of the data underlying the empirical parts of the project, and outlines the key challenge of how to empirically measure or categorise research topics, specialities, and disciplines. I explain how publications by individual grantees can form the basis of two different classifications of research content: A top-down classification aimed at situating a grant within a discipline or a disease area, and a bottom-up classification utilising publications' references to classify research topics. Chapter 5 gives a concise and brief overview of the core empirical results. The chapter cuts across the four empirical research articles by addressing the synthesis model of the dissertation. I show how funding is strongly concentrated across multiple conceptualisations of research content. I then discuss how prone funding distributions are to path dependency and positive feedback, and show that funding allocations tend to be very static but are not immune to change. I also present results of how funder priorities impact the distribution of funding in some contexts and at some levels of research content and at other times seem to only temporally shift these patterns. Chapter 6 discusses the results relative to the theoretical drivers out-lined below, discuss when funding concentration may or may not be a disadvantage, outlines some strong limitations of my results, while also offering some broader implications for funding policy and future research.

2. Conceptualising research content

Both the scholarly literature and the broader research policy debate are rife with speculations about how much, and why, competitive research funding is concentrated across different types of research content. Many of these beliefs rest firmly on the lived experiences of researchers and funding administrators themselves. Some argue that low success rates for grant applications and the conservative nature of peer review force researchers to propose research in well-established research areas or on topics with a proven record of results (Alberts et al., 2014; Stephan, 2012). Funding opportunities are seldom direct determinants of what researchers choose to study, but researchers are attentive to the organisational context of their choices, and adapt to them (Horta & Santos, 2020; Leisyte & Dee, 2012). Research effort and attention is also concentrated in a minority of research areas, disciplines, topics, and within the medical sciences, disease areas. Large-scale analyses of the scholarly literature show a clear core-scatter tendency. The bulk of research publications (namely journal articles) are published within a small core set of disciplines and topics while a large set of areas are more marginally addressed (Aagaard et al., 2019; Larivière et al., 2010; Waltman et al., 2019; Yao et al., 2015; Yegros-Yegros et al., 2020).

Why is that? Probably for a myriad of reasons, but an obvious expectation would be that funding and publication activity correspond closely (see e.g. Ciarli & Ràfols, 2019). If competitive research funding is indeed limited to a few areas of research or if projects are mainly proposed on tried and tested topics, the financial input to science could indeed be concentrated in a similar core-scatter fashion. But before we can answer such questions, we need to establish a clearer understanding of research content. Unlike most studies of funding concentration, which deal with the allocation of resources to individuals or institutions (Katz & Matter, 2020; Mongeon et al., 2016; Wahls, 2018b; Wahls, 2019), research content is not a straightforward object of analysis. In this chapter, I set out to clarify my notion of research content. A clearer distinction of different conceptualisations of research content is needed, not only to communicate what research content groups are, but also to connect these to questions of how concentration of research funding emerges. We can group research grants at different levels of aggregation by situating them within their broader research field or discipline, or the more specific specialties or topics addressed in a project. Patterns of funding concentration or dispersal can possibly be very different at different levels of aggregation. Funding is likely concentrated in the more cost-heavy research fields of biomedical and natural sciences, and certain disciplines such as experimental physics rely on more expensive equipment than do philosophy. But how is funding distributed within subdisciplinary areas with similar cost structures? To answer these questions, we need to categorise grants at many different levels of research content. Moreover, the question of why funding might be strongly concentrated or dispersed will likely have very different answers depending on the level of aggregation. Political priorities and funder strategies may play a prominent role in what research fields and disciplines to promote and strategically fund, while funding dynamics at the more granular levels of research specialties and topics are likely more removed from such influence, and instead reflect priorities of individual researchers and the incentive structures within their fields.

2.1 What I talk about, when I talk about research content

I use the term "research content" to signify a grouping or categorisation of related pieces of research, whether documents, articles, or grants. Grouping of documents into subjects or topics has been a hallmark endeavour in Information Science in order to foster easier information retrieval and to facilitate indexing of scientific texts. However, the notion and underlying theory on how scientific material is related and form a subject, topic, or indeed reflect similar content is not completely settled (Hjørland, 2001, p. 776).

In grouping different scientific documents, including grants, into topics or subject categories, we are essentially trying to discern what they are about, or their aboutness in the language of information science. Despite some disagreement on this part (see e.g. Hutchins, 1978; Hjørland, 2017), aboutness refers to our (the researcher's) effort in condensing the content of a scientific document into a subject category or index entry (Hutchins, 1978, p. 172; Hjørland, 2001). Discerning a summary topic or subject from a text is an intellectual indexing process involving a great deal of choice and subjectivity, because a subject categorisation rests on making judgements on certain properties of a document. While a document, such as a grant proposal, is the author's subjective treatment of a treated topic, it also has objective properties. We can imagine an almost infinite amount of document properties, such as a book's binding, its author, or the type setting. But often, a few objective properties are sufficient to group texts into subjects (Hjørland, 1992, pp. 182–183). One example is the relational properties of a document. How a grant proposal or journal article treats other documents through references, treatment and criticism of other scholarly work, tells us something about their shared subject or aboutness. Other properties could be language-based, including the use of certain words, the frequency of different words, or the structure of sentences.

According to Hjørland (1992, p. 183), classifying texts according to their subject is a two-fold process of assigning specific "predicates" to these texts. First-order predicates involve selecting appropriate properties of the document. These properties are the ways a document represents reality, and which parts of reality it represents is the aboutness (Hjørland, 1992, p. 182). A second order predicate then involves using these properties in the assignment of a subject category. Why may this be important? It highlights that grouping research into classifications of research content involves both a choice of which properties to enter into the subject classification, and how they are used in the classification task. While a small number of properties are often enough to designate a certain subject label, a choice among potentially infinite document properties constitutes a very subjective decision in categorising research. Consequently, how we relate grants or research publications are not neutral representations of the properties signalling what a piece of research is about and its content. Second-order predicates are not neutral, objective, and exhaustive representations of the document properties.

In the following, I outline a number of theoretical challenges we might face when categorising research according to their content. I also discuss three perspectives on categorisation, which highlight different relevant document properties used to facilitate large-scale research content categorisation: An institutional/sociological perspective, a cognitive/relational perspective, and a semantic/communicative perspective. Neither of these encompass all possible approaches to content categorisation. Instead, they are some of the most well-known and used frameworks

2.2 Challenges in research content categorisation

As categorisations based on the aboutness of a document are not a perfect representation of its research content, we may argue that such categories, at the surface level, represent its topicality. In this sense, topicality and aboutness are overlapping descriptions of research content, albeit not completely equivalent (see e.g. Hjørland, 2017, p. 61). But determining the content of a piece of research does not rest solely on determining its aboutness. Another important parameter is the relevance of a document.

Relevance and topicality are partly overlapping concepts (Morris & Van der Veer Martens, 2008, p. 226), but the relevance of a document concerns the usefulness of this document for some specific user need. In the literature on indexing and information retrieval, the relevance of a document partly relies on its match with a given topic a user is interested in, and the "(…) appropriateness of that topic has to do with whether or not a piece of information is on a subject that has some topical bearing on the information need in question." (Greisdorf & O'Connor, 2003, p. 1296). A document is then relevant for a topic if it constitutes the best match for the topic of a query and has informative potential for a task (Mizzaro, 1997, p. 812; Schamber et al., 1990, p. 758; Hjørland, 2001, p. 777). However, relevance of a document is also a distinct part of its content, not always aligned with its topicality. A central premise is that a content categorisation represents both the content meaning of a document and the information need of a user (Schamber et al., 1990), but we could easily think of a situation where a document is on topic but not relevant. A journal article on schizophrenia from a psychological perspective may not be relevant to a psychiatrist researching certain pharmaceuticals, even if it is concerned with schizophrenia (the topic). Therefore, the same topic grouping may appear more or less relevant to different observers, depending on their specific needs and views. Why is this a problem for studying how funding is distributed across different research topics, fields, etc.?

Firstly, the categorisation of a large number of research grants, based on certain properties of these grants, will not yield completely accurate content categories for all observers. A categorisation is not a priori exhaustive. The relevance of individual content grouping will vary, because different properties used for categorisation (i.e. first-order predicates) are more or less relevant to different researchers, policy makers, or other users. This problem of varying relevance imposes some important constraints on how much we can generalise from a broad study of funding distributions. Identifying important properties for categorising research grants from the narrow perspective of individual researchers could potentially ensure both high topicality and relevance and is more pragmatic (Hjørland, 1992, p. 183), but is also more problematic because of a limited scope. I am interested in describing broader trends across many research fields, and not just a few selected subjects, which necessitates a broader approach: Categorising a large number of grants across multiple disciplines and topics. While this facilitates broad comparisons of funding distributions, it comes at a steep cost of low precision. Why is that? Because categorising many grants across varying fields of research can only be accomplished by using a small set of "document properties" (i.e., metadata about the grants) to conduct the categorisation. The problem then arises because the "objectivity" of a topic categorisation rests on selecting a few properties acknowledged by many to signal something about research content. However, using very general properties may be inversely related to the precision or "objectivity" of the topic categorisation. Only certain researchers, with strong contextual knowledge of a topic, have a complete grasp of the significant properties of a grant. They know the relevant keywords and the important references situating a grant within a topical area, while properties identified by many are sure to be much less descriptive of the research content (Hjørland, 1992, pp. 182–183). In sum, when categorising many grants across many topics, we likely end up with a non-exhaustive categorisation. This would not necessarily be relevant to individual researchers interested in how much funding is spent on the two or three topics they work with. It is an inherent trade-off when automatically indexing research. We cannot with strong precision pinpoint which exact topics are highly funded and which are not, but are confined to examining the broader trends and distributions of funding allocations.

Categorisation is further complicated by a second challenge. Our ability to discern the subject of a piece of research rests on both epistemological and methodological quicksand. In the eyes of individual researchers, the grouping of research topics or content may not resonate completely with their notion of a certain grouping of research (Held et al., 2021). But in the aggregate, the grouping of research topics or content may provide an expedient approach to assessing the spread of funding across different types of research. However, we have very limited ways of assessing how precise or "objective" our categorisations are as any "ground truths" about what constitutes a specific topic, discipline, or research field is often lacking (Gläser et al., 2017, p. 983; Klavans & Boyack, 2017c, p. 984; Klavans & Boyack, 2017a, p. 1160).

The establishment of an objective partitioning and categorisation of research (i.e., a ground truth about subject categories) is hampered on both epistemological and methodological grounds. In an epistemological sense, we could think of two pertinent perspectives: Subjects are either objective but not clearly identifiable, or subjects are subjective and an interpretation by the user of a classification. The latter entails that there is no absolute notion of categorisation accuracy and each attempt to group or cluster research content will be accurate/inaccurate in its own right. Discussing the most objective way to group and label different grants according to their aboutness would then be futile. Two schemes for categorising grants, whether based on their textual properties or what research they cite, may then highlight equally valid aspects of their topical differences and similarities, and provide perhaps differing or similar limited perspectives on the concentration of funding (Waltman et al., 2020, pp. 691–692). The former instead assumes that some authoritative partitioning of research content exists, but cannot be observed. According to this, some categorisation schemes are more accurate than others, but it would take an impossibly large host of domain experts to asses which approach yields a more accurate subject classification (Waltman et al., 2020, pp. 691–692).

Recent research has argued for comparing automatic, algorithmic, classifications of science literature with assessments from the authors who wrote the publications in question (Held & Velden, 2019; Held et al., 2021). Interviews with authors reveal little consistency between automatically constructed topics and the perceived topics of the authors themselves (Held et al., 2021). Others have used existing classification schemes, based on curated and manually labelled Medical Subject Headings, and examine how data-driven automatic indexation fits these existing categories (Sjögårde et al., 2021). Neither method of validation appear wholly convincing. While authors are arguably the authority on the aboutness or topicality of their publications, the discrepancies between their understanding of a topic and a categorisation may depend on the relevance they attach to publications grouped together with their research. Researchers working on the same or similar topics are not even in agreement on what methods were or were not integral to their topic (Held et al., 2021, p. 4518), casting doubt on the notion that author judgements are clear paths to establishing ground truths. As for the comparison between existing and automatic classifications, it is not clear how discrepancies are to be interpreted. Are differences between curated vocabularies and automatic classifications based on keyword or citation properties a reflection of severe shortcomings on the part of the automatic or the existing classification (Waltman et al., 2020, p. 692)? In each instance, the lack of an established benchmark is a major challenge to evaluating how well we can measure or construct research content classifications.

In the context of this dissertation, I rely on a variety of different classification schemes when grouping grants (elaborated below and in chapter 3). It is important to remember that these do not imply some authoritative landscape on what type of research is being funded. Instead, they are merely one view of the distribution of funding and funding portfolios. It is thus important to remember when interpreting results, that categories are not neutral or independent of the methods used to construct them or the interpretation of the classifications of research content. When categorising, we assume that a given procedure can "(…) penetrate the surface of documents and reveal the hidden subject." (Hjørland, 1992, p. 179). The operationalisations presented in chapter 3 all rely on a limited number of properties connected to individual research grants, and mostly properties of research articles published by the grant holder. The subject classifications I employ are then best described as indirect and non-exhaustive representations of the research content in each individual grant.

Lastly, I want to highlight an important set of interrelated challenges pertaining to the hierarchical nature of research content classifications. As argued above, we can levy different properties of a document to classify its contents. More general properties will often be the most widely recognised by a wide selection of users, but also often the least descriptive ones. A research article published in the journal ACS Nano is easily recognisable as dealing with research in nanotechnology, but if we are interested in research on graphene or carbon nanotubes, this classification is not specific enough. The hierarchical ordering of subjects are of course not endemic to the categorisation of research content. Most taxonomies, from classifications of political regime types to phylogenetic trees, are ordered with different levels of abstraction. This is natural and desirable as it allows us to move between broader or narrower views of a document's aboutness.

Here, I address the distribution of funding across several levels of subject classification; from the broad research fields and disciplines to more specific topics or disease areas. Figure 2.1 illustrates the hierarchical nature of the research content classifications I employ. Most of the empirical analyses focus on the "middle" parts of the classification hierarchy. In Paper B and D, I investigate funding concentration from the perspective of disciplines to show that not only do the physical and biomedical sciences receive the largest share of funding allocations, but within these research fields, certain disciplines such as molecular biology and clinical oncology are far ahead of e.g. plant sciences in the competition for funding. As elaborated below, disciplinary categories and labels are content classifications often derived from the institutional and organisational setting of the conducted research (Hammarfelt, 2020), such as the department of the authors or the publication outlet. Paper A, B, C, and D venture further down the ladder of granularity by also focusing on research specialities and topics. In the articles, I primarily use the terms "meso topic" and "micro topic" to highlight that the delineation of topics vs. specialities are not always straightforward.

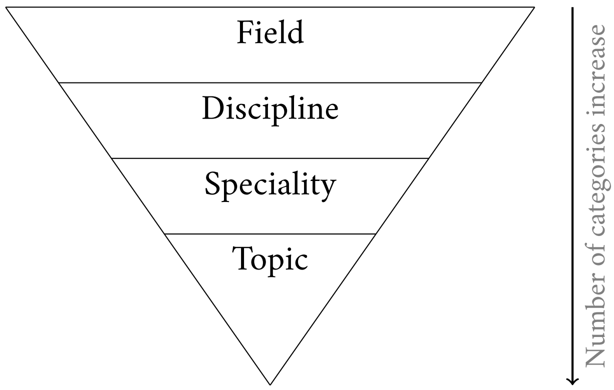


Figure 2.1: Hierarchy of research content levels. Terms used for categories at different levels for research content classification. Research fields are assumed to be the broadest, less descriptive, categories, and each step down adds more categories, detail, and granularity.

A speciality is a self-organised social group defined by a common focus on one or a few research topics, a cohesive core of researchers, and a shared degree of basic knowledge and vocabulary (Morris & Van der Veer Martens, 2008, pp. 219, 233). The members of a speciality are more frequently in communication with each other than with other scientists and share a consensual structure of concepts through citation of each other's work (Small, 1980, p. 183; Morris & Van der Veer Martens, 2008, p. 220).

Early accounts suggested that these specialities consisted of around 100 individual researchers (Kuhn, 1970, p. 178; Price, 1986, p. 64). For different reasons, the meso topics I use likely consist of many more researchers. I make use of topic classifications consisting of around 800 (meso) and 3000 (micro) different categories, which often cover more than 100 researchers' cumulative work. Compared to other classifications of research topics (Boyack et al., 2011; Boyack & Klavans, 2005; Boyack et al., 2014; Klavans & Boyack, 2011), these are not very fine-grained or granular. Another reason is the general development of the scientific enterprise, which has expanded greatly since the 1970s and 80s in terms of larger teams, more researchers, and many more publications (Bornmann & Mutz, 2015; Milojevic, 2014; Price, 1965; Wuchty et al., 2007). Research topics are perhaps similar to research specialities in that they are made up of a selection of documents (and their authors) with a common intellectual focus, such as research problem, and mutually refer to each other due to their shared research questions (this definition is similar to Collins, 1985; Klavans & Boyack, 2017b, p. 4; Klavans & Boyack, 2017a, p. 1159). A topic can therefore be thought of as the subject often addressed in review articles that summarise the state and development within a given problem area (Klavans & Boyack, 2011; Sjögårde & Ahlgren, 2018, pp. 135–137). In some areas of science, a speciality may only encompass a single topic, while other fields may have specialities made of more topics.

This is also the gist of the challenge from hierarchical ordering in science studies: When hierarchically ordering research content, we will inevitably end up with a core and scatter pattern. Some documents will be at the core of a given subject classification, while a large number of documents will be more tangent to the label or grouping we have constructed. In essence, each constructed topic or speciality will be made up of a heterogeneous set of e.g. research grants or publications, which can be classified into even smaller groups. This core-scatter phenomenon arises because the items studied in information science, scientometrics, and science studies are already subject to very skewed distributions (Morris & Van der Veer Martens, 2008, p. 237; Price, 1976). A core set of authors produce the majority of research articles, while a large scatter of researchers produce only a small percentage of all scientific literature (Lotka, 1926). Similarly, a small core of journals publish the majority of work within a given subject, such that one journal in this group encompasses as many relevant articles in n journals in the next group, which then publish as many articles as n^2 journals in the following group (Hjørland & Nicolaisen, 2005, p. 97).

Core and scatter patterns produce at least two problems for the classification of research grants. For one, the classification may primarily embody the core elements (grants) of a topic or speciality. Clusters of topics and specialities with high degree of funding may not necessarily represent a strong concentration of funding on all types of research within that topic or speciality. The core-scatter patterns of most classifications could obscure the fact that some research problems within a speciality (those forming parts of the scatter) will not be very well funded while problems addressed in the core may be.

Next, the core and scatter structure of subject classifications also lead to some overlap between specialities not reflected in the categorisation (Morris & Van der Veer Martens, 2008, p. 238). Some grants and publications partly related to multiple subject categories will be placed in one or the other category, even if speaking to the same type of research content. This could both artificially inflate or underestimate patterns of funding concentration. If we assume colon cancer to be topical (if not a topic), we can hypothetically imagine a situation where research on 'treatment of colon cancer' and research on 'rehabilitation after colon cancer' are grouped separately. It may be that the former is grouped with related research on surgical or pharmaceutical treatment, while the latter is grouped among research into post-treatment relapse prevention. If many grants address the treatment-related problems, we could be led to conclude that research into colon cancer is much more well-funded, while some parts of this speciality are not. Conversely, if colon cancer rehabilitation is grouped with e.g. rehabilitation after breast cancer, we may miss that not all grants classified with this category is concentrated among breast cancer research.

All together, these challenges provide some important barriers towards categorising research grants into fields, disciplines, specialities, or topics. Creating a satisfactory, exhaustive, and accurate classification of the type of research conducted within a funded project becomes nearly impossible. A lack of ground truths about topics or specialities hinders a thorough evaluation of any classification's accuracy. Varying relevance or what document properties are most descriptive in a given case will leave some classifications unsatisfactory to individual researchers, and overlaps in subject categories hinder exhaustive and mutually exclusive topic classifications. These barriers leave this dissertation at an impasse, equally frustrating and exciting. In order to classify grants to analyse the dispersal or concentration of funding, certain properties need to be selected for this undertaking. They will necessarily result in imperfect attempts to "measure" research content at a large-scale, quantitative, level. However, the difficulty of this undertaking also necessitates experimentation with different approaches. In the following section, I outline three theoretical perspectives on aboutness or topicality, which highlight different properties of documents as descriptive of their content, and argue, for my use of two of these perspectives.

2.3 Perspectives on aboutness and topicality

To reiterate the basic tenants of the discussion above: I use the term ”topic” to signify the lowest level of aggregation in terms of a thematic structure, ”speciality” or ”meso topic” as a slightly broader classification of content, which could encompass more than one topic, ”discipline” to denote a broader, institutionalised field of research that covers different topics, and ”research field” as a collection of disciplines such as “social sciences” or “physical sciences”.

These levels of analysis can be conceptualised differently depending on how we choose to view the structure of science, or more precisely, the properties we use to demarcate the different subject categories from each other. I focus specifically on three approaches to the conceptualisation of research content categories: An institutional or sociological approach, a cognitive/relational approach, and a semantic/discursive approach. These are not exclusively the only approaches, they consist of different sub-approaches, and are frequently named differently (see Morris & Van der Veer Martens, 2008, pp. 220–235). However, they are common ways of both thinking about topical structures, each point to distinct types of relevant document properties used for categorisation, and offer different ways to operationalise research content categories when applied to data.

In an institutional or sociological approach to aboutness and topicality, the social properties of documents are highlighted in an effort to group research according to the underlying social, institutional, and organisational context in which they originated or came to be. The sociological approach to studying the organisation of science was once dominating (Chubin, 1976, p. 448), and largely followed a Mertonian, functionalist, view of how research specialities and disciplines are ordered (Morris & Van der Veer Martens, 2008, p. 221). The specialisation of science can be viewed as a function of its social ordering, where cognitive development (e.g. the research content examined) developed as a function of how a speciality or discipline was organised (Morris & Van der Veer Martens, 2008, p. 222). According to Merton (1957; 1968) and others, the functioning of science operates on strong principles of priority for scientific discovery and a stratified system of recognition based on the scientific merit of these discoveries. Differential awards create various strata or hierarchies among scientists, where highly placed researchers can legitimately claim cognitive superiority for certain ideas and topics, thereby directing the content examined within a speciality or discipline. As summarised by J. R. Cole and Zuckerman (1975, p. 143): “(…) development and elaboration of the cognitive structure of new specialties appear to depend in part on correlative development of their social structures—on the routinization of an evaluation and reward system, procedures of communication, acquisition of resources and the socialization of new recruits.”

What does this imply for the relevant properties of a grant or document in terms of categorisation? Because evaluation systems, resource allocation, and teaching (i.e., the socialisation of “recruits”) noticeably differ between different university departments, research groups, or even academic journals (think of peer review), this approach assumes that institutional properties of a document are important for describing its content. Where the author is employed and the publishing outlet then becomes central properties for categorising. Various categories are then demarked by different university departments, which have been tasked with teaching and training university students and the next generation of scientist in the methods and theories of a given discipline or speciality, and become instrumental in defining its content (Chubin, 1976, p. 448). From a research funding perspective, funders allocate resources to departments, laboratories, or research groups, and the group of people within these units become the embodiment of a discipline or speciality (Hagstrom, 1970, p. 93; Martin & Irvine, 1985, pp. 559–560). Similarly, research can be categorised into disciplines according to the type of scientific journal or conference it appears in. This type of categorisation rests on the assumption that the scientific literature is specialised, and scholarly journals provide outlets for each specialised area. Just as organisational units, such as departments, journals and conferences become institutional boundaries (Klavans & Boyack, 2017b, p. 4) that divide research by sorting what is relevant and what is not, and can serve as anchors for communication within a scientific area (Milojević, 2020, p. 184).

However, an institutional view of aboutness also misses the mark on several points. By categorising research based on author affiliations and associated publication outlets, social properties end up "black boxing" the actual content by conflating e.g. departmental labels with the substances of a text (Whitley, 1970; Chubin, 1985, p. 232). It also ignores the underlying heterogeneity in content. Journal or departmental labels are often not detailed enough to describe their differential research output. Journals may embody a range of different types of research transcending disciplines and specialities (Boyack & Klavans, 2005; Leydesdorff & Bornmann, 2016). Moreover, institutional structures are highly stable. Journals, conferences, and university departments seldom die out, and using these to define a thematic structure in science will gloss over some underlying changes (Boyack et al., 2014, p. 150; Price, 1965, p. 515). The type of research conducted at a department could change significantly over time, and journals or conferences may shift focus to related but distinct research content while retaining their name. Furthermore, the Mertonian approach assumes social structures to be the antecedents of the cognitive development within a speciality or discipline, whereby reward systems and teaching organisations give rise to and facilitate convergence onto similar research problems. It could equally likely be that intellectual or cognitive developments precede and influence a later social structure (Chubin, 1976, p. 449).

Given these challenges in using institutional labels as properties to categorise research content, we should be sceptical about attempts to classify topics or specialities in this way. As argued by Morris and Martens (2008, p. 239), departments, journals, and other institutional entities are not homogenous enough for detailed mapping onto specific research specialities. But perhaps we can use the sociological and institutional approach to instead classify grants into higher rungs of the hierarchy proposed in figure 2.1 above, such as disciplines?

A central tenant for disciplinary structures are their institutional and organisational roots. What distinguishes disciplines from content classes lower in the hierarchy are their explicit connections to departments, journals and conferences (Hammarfelt, 2020, pp. 245–246). Disciplines should not be confused with topics or specialities because they are knowledge institutions in themselves, embedded in structures for teaching and disseminating knowledge on a specific object of research (Krishnan, 2009; Hammarfelt, 2020). The establishments of disciplines is specifically rooted in the control over channels for disseminating knowledge relevant to the object of research, such as journals, book series, and conferences (Hammarfelt, 2020, p. 246). As argued by Sugimoto and Weingart (2015, pp. 778–779), the definition of a discipline have often rested on an institutional component, including academic departmentalisation, the establishment of degree-granting departments in many universities, and the development of discipline-specific academic journals. The establishment of journals can serve as one central narrative embodying specific disciplines. Examples include the founding of the Journal of Evolutionary Biology, or the American Journal of Sociology as key events creating a disciplinary structure (Sugimoto & Weingart, 2015, p. 784).

In this dissertation, I use content categories based on journal articles to label research grants, by connecting the articles of a grant holder with the grant. These articles become the basis of disciplinary categories. Chapter 4 provides a more technical introduction to the disciplinary categories. Using Web of Science's and other journal-based subject categories to gauge sub-disciplinary, disciplinary and interdisciplinary tendencies in science has been widespread (e.g. Larivière et al., 2016; Gazni et al., 2012; Ràfols & Meyer, 2010; Ràfols et al., 2010; Ràfols et al., 2012; Rotolo et al., 2015, p. 225; Milojević, 2010; Milojević, 2020; Sugimoto et al., 2019). However, this approach is not without drawbacks. First, the institutional basis of disciplines (such as journals) are only partly relevant properties. These types of properties fulfill one aspect of what a discipline is: A group of persistent and stable institutions across time (Sugimoto & Weingart, 2015, p. 781). Relying on rigid, stable and predefined categories runs the risk of ignoring emerging or dynamic changes to disciplines (Ràfols & Meyer, 2010, p. 269). Furthermore, it overlooks the cognitive and communicative aspects of disciplines. Most definitions of disciplines are composites, highlighting that disciplines are "groups of researchers working on a specific set of research questions, using the same set of methods and a shared approach" (van den Besselaar & Heimeriks, 2001, p. 2) or a "body of knowledge and a social body that generates, evaluates, communicates, and teaches the corresponding knowledge" (Schummer, 2004, p. 436) (both quoted in Sugimoto & Weingart, 2015).

These shortcomings point towards other approaches towards measuring aboutness or topicality.

A second approach emphasises the cognitive or relational similarity between documents as important properties for classifying them. Before I delve into this approach, a disclaimer is warranted. In discussing the cognitive/relational approach to research content classification, I am painting with a broad brush, and perhaps describe different approaches in an overly generalised fashion. How to classify text into scientific content categories is an ever expanding area of study, and even if important to the work presented here, somewhat tangent to the broader aim of this dissertation.

The cognitive/relational approach to classification of research content broadly assumes that researchers working on similar problems and topics will mutually refer to each other or to the same third party work (Morris & Van der Veer Martens, 2008, pp. 224–227). Much of this work assumes that citations between e.g. research articles is a manifestation of a communicative relationship between the citer and the cited. From this view, a reference is presumed to signal that the cited work was important to work citing it (Chubin, 1976, pp. 451–452). A focus on direct citations between research articles, or co-citations (where two documents are cited in the same work) has been integral to previous mappings of research specialities or topics. For example, Chubin (1976) suggested that the nature of communication relationships between researchers are integral to the formation of specialities. If we return to the definition of specialities and topics above, a central part is the view of specialities as cognitive fields populated with documents referring to each other. We should then be able to trace backwards from citation relationships to classify mutually citing documents as having similar research content (see also Small & Griffith, 1974). This approach rests on the assumption of normative citing (Nicolaisen, 2007), whereby citations are a primary way of recognising the work of others, and a way to acknowledge intellectual debt. Grouping research content based on citations presumes that citations reflect the merit of what is cited, is content-wise related to the citing document, and that all citations carry equal weight (Nicolaisen, 2007, p. 615).

Compared to the institutional approach, a cognitive/relational view of aboutness leaves more room for change within a topical structure. A topic can disappear or change once the next theory, method, or key advance comes along (Boyack et al., 2014, p. 150), because authors start and stop referencing certain documents. Similarly, a cognitive view of topics allow for the composition of a topic to change when new authors enter, former members change referencing behaviour, or some authors are no longer being referenced (Chubin, 1976, p. 450). Topics are therefore dynamic as they include groups of scholars specializing in and writing about the same themes, referencing articles they consider relevant, regardless of whether they agree or disagree with those articles

(Bruggeman et al., 2012). Lastly, a cognitive definition frees thematic structures form their institutional environment. A topic can then be addressed by many different research environments, communicated through different journals, and different journals communicate work across different topics. It encompasses a smaller set of units, the documents and authors with common referencing behaviour, and can be applied to demark between more precise topics (Hagstrom, 1970, pp. 91–92) instead of disciplines. This is also my approach to discerning topics and specialities (or micro and meso topics). As outlined in chapter 3, I used articles published by grant holders and relate these to other research articles through their direct citations, and superimpose the articles onto the grants themselves. As with my approach to classifying the disciplinary category of individual grants, this is by no means unproblematic.

Using citation relationships to operationalise the aboutness of research grants is surely a crude and only partly satisfactory solution. Citation linkages between papers are only an approximate way of measuring intellectual debt (Morris & Van der Veer Martens, 2008, p. 227; Chubin, 1976, pp. 451–452). Particularly, the assumptions of normative citation behaviour are likely to be flawed. Mentioning of a previous document is not necessarily an objective indication of influence or content similarity, simply because the motivations for citing are not necessarily merit-based. Studies of citation motivations highlight that citing other research is primarily an attempt to persuade rather than give credit (M. H. MacRoberts & MacRoberts, 1986; M. H. MacRoberts & MacRoberts, 1987; M. H. MacRoberts & MacRoberts, 2010). Many citations are given, not to reflect influence of other research, but to distort the evidence to fit the authors own arguments (Nicolaisen, 2007, p. 617). Authors are also influenced by the perceived worth of the work they have available to them, and may not cite the most relevant or cite work irrelevant of the context (Nicolaisen, 2007, p. 616). Moreover, citing may also be unconsciously influenced by non-merit based factors, such as prestige or previous citations. Often, we cannot clarify why a reference is deemed relevant, because we fail to consciously think of reasons for citing a particular source and not citing another (Nicolaisen, 2007, p. 615). Because previous citations may influence the motives for citing an article, this likely creates core and scatter patterns as argued above. As core and scatter phenomena produce skewed distributions of references with few sources being cited much more than the rest, it becomes harder to distinguish groups of entities (Morris & Van der Veer Martens, 2008, p. 239).

Lastly, we can also approach the categorisation of research topics from a semantic or discursive approach. A semantic view of aboutness eschews the common referencing behaviour across texts, and instead emphasises the importance of shared word use across texts with similar content. In this sense, a semantic-based categorisation of research grants place more emphasis on the actual content of individual project descriptions (Griffiths & Steyvers, 2004). This type of categorisation scheme have seen wide use in both bibliometrics, information science, and also computer science. One example is co-word analysis, where groups of documents are categorised based on their shared use of specific technical terms know to be central in a given speciality or topic (Morris & Van der Veer Martens, 2008, p. 229). Using shared keywords between documents as properties for categorising research seems especially helpful when mapping the priorities for different research within a speciality. Ciarli and Rafols (2019, p. 954) use both key phrases extracted from publication titles and abstracts, along with descriptors assigned to publications by trained librarians, to classify different topics within rice-related research. The problem with using only natural language from scholarly documents is the often near synonymous use of similar terms across topics. Terms extracted from title or abstracts may not accurately discriminate between different semantic or discourse communities, which makes them less optimal as sole indexing units. Descriptors from a controlled vocabulary or thesaurus instead bear the distinct advantage of capturing content and aspects of documents not reflected in e.g. titles (Ciarli & Ràfols, 2019, p. 954; Boyack et al., 2011), but also presuppose that some classification exercise has already taken place. In both instances, a topic is assumed to be a collection of documents with a shared organisation of words that represent an underlying semantic theme.

One way of bypassing the problems of shared term or vocabulary use across topics is to partially classify all documents within all topics, based on the frequency of terms, descriptive of different topics, used. This has been commonplace within computer science and the use of latent dirichlet allocation for topic modelling. Here, a semantic theme can be informally thought of as a certain organisation of words or a probability distribution over different terms in a vocabulary (Yan, 2014; Yan et al., 2012). Each document then reflects topics with a different proportion such that some topics are more prevalent within a text than others (Blei & Lafferty, 2006; Blei & Lafferty, 2007). However, we are then left with a classification system, where each document is not unequivocally related to one topic. While this likely reflects the actual content of a document, it is also somewhat self-defeating: Classification of documents into topics or specialities is meant to be complexity reducing, but topic modelling does not always yield a clear topic for a document. Furthermore, if some terms are frequently used across topics, most texts within a corpora will seem to partially consist of a few particular topics, and these topics are likely not very descriptive.

On the other hand, a semantic approach can be an advantage compared to a cognitive definition as researchers working on a topic may only implicitly build on existing knowledge from documents that are not the most relevant or recent (Gerow et al., 2018). For example, a researcher may build on a theorem derived in the 1950s, but only reference work, which elaborated on this in the 1980s. Furthermore, the primary artefact of scholarly work is not the context where it is used but the text itself (Gerow et al., 2018).

As an attempt to overcome shortcomings of both the cognitive/relational and semantic/discursive approaches, hybrid approaches may provide a good middle ground for detecting topics and specialties in publication or grant data. Some efforts have used citation relations as clustering components for grouping publications, and terms extracted from full texts as descriptors of these groups (Sjögårde & Ahlgren, 2020). Other approaches venture further by using both the citation relations and text similarity to classify research, in order to enhance the accuracy of one method with another (Ahlgren et al., 2020).

In the following, I make use of both the institutional approach to classify grants into disciplines, and methods based on citation relations and on one occasion textual descriptors from the Medical Subject Headings vocabulary to classify grants into specialities (meso-topics) and topics (micro-topics). Echoing Gläser, Glänzel, and Scharnhorst (Gläser et al., 2017, p. 986), topical structures likely shape researchers' work, but there are no "ground truths" about scientific topics. All three definitions capture relevant aspects, and entail different ways to measure and construct topic categories. In chapter 3, I lay out the methodological considerations for using each approach. It is important to fully acknowledge that neither will provide a complete, "right", picture of how funding is distributed across topics and disciplines. The general tendency for a core and scatter pattern, inherent in topical detection, would likely never produce evidence of equally distributed funding patterns. But hopefully, the approaches taken here provide some measure of empirical description of what is funded and how funding is allocated across categories of research content.

3. Theoretical perspectives on funding concentration

This chapter outlines a theoretical framework for analysing research funding concentrations across both individual researchers and research content. The chapter is structured into three parts: The first part outlines some broad considerations of what structures the supply and demand of different research content. Besides describing patterns of concentration and dispersal, we need to consider how research supply and demand is created, what influences it, and how funding allocations may play a role. In this part, I focus especially on one of the most basic perspectives on the supply and demand for research: The linear model of science. It stipulates that the type of research supplied is primarily a function of what researchers themselves consider opportune and worthy of further study. From this perspective, demand for particular types of research is created *internally* in the scientific system, and ensures an optimal and effective allocation of resources to different problems.

However, I also outline a strong challenge to this perspective, and stress the importance of some key institutions of reward and recognition in the science system: peer-review and citations. I argue that even if the supply of science is mostly affected by what researchers demand, the use of peer review in allocating funding will change the incentives, and correspondingly the demand, for studying particular topics or work within certain specialities and disciplines. Based on a sociological perspective on science, the uneven distribution of recognition shapes what type of research is considered worthy of funding. In the second part, I argue for the importance of considering priorities from actors external to the system of science, namely the funders sponsoring much of what takes place within science. As competitive funding has taken up a growing share of university research resources, funders are increasingly placing demands on what they are willing to fund, both directly through earmarking and thematic targeting, but also indirectly through their contribution to the creation of strongholds within certain fields.

Finally, I bring these three strands together. My main argument is here that external and internal mechanisms are hard to clearly separate and the relationships between them are characterised by complex interactions. I illustrate this by suggesting a simplified model of how internal and external factors can interact to create both positive and negative feedback in the funding system. The concentration of funding is a result of a complex interplay between internal and external factors shaping what composition of research is demanded and supplied, which may create more or less funding concentration at different times. Based on this, I derive some important problems for empirical analysis and discussion.

3.1 Supply and internal demand for science

Assessing the link between competitive funding arrangements, the individual choice of what to study as well as the broader distribution of what is studied in the scientific system is an almost herculean task. The direction of science is at least to some extent influenced by research councils and other research funders, but how this influence plays out, and how specific changes of research content take place is not well understood (Gläser & Laudel, 2016, p. 123). Furthering our understanding of this is complicated. Researchers themselves are the endpoint when it comes to choosing what topic to study, and are not simply passive recipients of funding decisions and policy (Laudel & Gläser, 2014; Leisyte et al., 2010; Luukkonen & Thomas, 2016). Because most researchers work within highly specialised fields, there are limits to external influence from policymakers, funders, and consumers of research because they are unfamiliar with the instruments, methods, and pressing research questions in these fields (Braun, 1998, p. 808). Therefore, competitive research funding arrangements do not dictate research content, but nevertheless serve as important scope conditions for the individual researcher's decision of what to study, and for the collective focus in the scientific community.

The question of how funding (and researchers' attention) is allocated across different problem fields has been framed as a question of how the supply and demand functions of science interact. In this view, decisions of individual scientists and science policy makers interact to shape the composition, allocation, and size of the research portfolios, which in turn supplies scientific results. Consumers of scientific results, whether researchers or external actors using scientific inputs, constitute a type of demand function influencing what scientific results are produced (Bozeman & Sarewitz, 2011; Sarewitz & Pielke, 2007). What type of research is funded may then be said to rest on an intricate interaction loop of supply and demand functions, and the ways in which funding arrangements constrain or shape these functions. Here we may inject that the concepts of supply and demand are overly simplistic. What type of research that end up being funded is a complex process influenced by a diverse set individuals and institutions, each with their own incentives, agendas, and capabilities (Sarewitz & Pielke, 2007, p. 6). But that does not negate the actual allocation decisions being made by researchers and funders alike. So while thinking of these allocations as an interplay between the supply and demand functions obscures the complexity somewhat, this heuristic can help in illuminating some of the underlying drivers of funding portfolios and the degree of concentration or dispersal of these.

3.2 Science as a self-regulating system

One perspective aiming to understand the supply of science and how attention is divided among different topics, specialties, and disciplines views the scientific system as inherently functionalistic and self-regulating. This view emphasises the role of the suppliers of science as driving the allocation of attention to different problems. The individual researcher has to choose between different topics, the research director which projects to push for, and the funding administrators which efforts to allocate resources to. The totality of these choices would then make up science policy as a whole, and the distribution of scientific effort (Carter, 1963, p. 172; Weinberg, 1963, p. 159). In this perspective, the central problem is a problem of problem choice: "In a world of finite resources, how should policy-makers choose among the many competing scientific disciplines, projects, and programs in making public investments?" (Bozeman & Sarewitz, 2011, p. 2).

According to at least two perspectives, problem choice in the scientific community is guided by self-regulation and self-coordination on the part of the researcher. Individual scientists choose to pursue projects and topics dependent on what other scientists have deemed worthy and relevant. According to Polanyi (1962), topics or problems are judged on their potential to lead to results that are (i) plausible, (ii) have scientific value (including importance for a research field and intrinsic interest), and (iii) are original. Each scientist is then assumed to pursue topics optimising these three criteria. Firstly, a researcher will search for topics that are likely to lead to plausible results in order to eliminate scientifically unsound problems or problems that are incongruent with current scientific knowledge and opinion. Second, she would weigh a topic's propensity for accuracy, importance, and intrinsic interest. Polanyi himself offers the choice between physics and biology as an example of this weighing. While the study of inanimate objects in physics lacks the importance of biology's focus on the living, it makes up for this by being extraordinarily accurate in its predictions and theory and vice versa (Polanyi, 1962, p. 58). Lastly, a researcher will prefer a topic with a sufficient probability of surprising and original results. This last criterion is simply an expectation that researchers will seek out problem areas where they are most likely to find unexpected, yet plausible discoveries.

Together, this individual weighing by each scientist leads to a self-regulating, scientific, market of ideas organised to identify and pursue the most efficient lines of research (Polanyi, 1962, pp. 61–62; Sarewitz & Pielke, 2007, p. 7). By Polanyi's own account, the distribution of funding would follow the same principles as long as allocations were based on scientific guidance, or more precisely, on peer review by other scientists (Polanyi, 1962, p. 61). This perspective describes an "ideal-typical" process of topic choice for both the individual researcher and a description of the allocation of attention in the scientific system as a whole. It is a rationalist model of topic concentration because it assumes that the supply of scientific knowledge is generated without any connection or attention to demand for particular types of knowledge (Sarewitz & Pielke, 2007). Instead, the spontaneous self-coordination of individual scientist effectively allocates attention to the most important topics, and topical or disciplinary concentration is a function of differences in plausibility, value, and originality.

A similar perspective was espoused by Vannevar Bush in "Science, the Endless Frontier" (Bush, 1945), which stipulated a model for scientific priority setting later dubbed the "linear model of innovation" or the "science push model". In Bush's view, scientific progress is of vital concern to governments because of its downstream impact on national health, the standard of living, and job creation (Bozeman, 2020). While these are different criteria than those employed by Polanyi, Bush's main thesis was that investing heavily in basic research would provide a reservoir of broadly applicable knowledge, which society could draw on to solve a multitude of problems (Sarewitz & Pielke, 2007, p. 7). Scientific supply should still rest on a mechanism of self-regulation, where strategic investment choices between topics and disciplines should follow from autonomous scientists guided by the logic of nature (Sarewitz & Pielke, 2007, p. 7).

In summary, the view of science as a self-regulating system sees funding concentration as a function of differences in plausibility, scientific value, and originality among topics, specialities, and disciplines. Varying plausibility and value is likely to create conformity in what is researched, while the value attached to originality leads to change or dissent (Polanyi, 1962, p. 58). The self-regulation perspective also prescribes a more normative guidance of how to arrange the funding system. Both Bush and Polanyi were ardent defenders of a funding system based on curiosity: "Scientific progress on a broad front results from the free play of free intellects, working on subjects of their own choice, in the manner dictated by their curiosity for exploration of the unknown" (Bush, 1945, p. 10). Outside demands and efforts to guide the direction of science was seen as an interference with its advancement (Sarewitz, 2016). The competition for grants and funding is then a functional mechanism for optimising scientific effort, as scientists will compete to solve the most pressing and important problems first (Hagstrom, 1974, p. 7). In a theoretical perspective, the expectation will be that floor funding for universities and *responsive mode* grants will support this effort, while thematically earmarked funds may disturb the balance and lead to sub-optimal allocations.

3.3 Institutional constraints on self-regulation

From another internal perspective, sociologists of science were quick to question the idea of optimal self-regulation (Allison & Stewart, 1974; Hagstrom, 1974; Merton, 1968; Zuckerman, 1967; Zuckerman, 1970). While the scientific system may operate on a functional basis, it is not only the informed weighting of plausibility, value, and originality that leads to a particular allocation of effort and attention. Instead, internal institutions of reward and recognition are important for understanding the concentration and dispersal of funding.

Institutional arrangements for distributing rewards and recognition are widespread in the scientific system. The most well-established of these are the norm of subjecting research to peer review, where other scientists judge the merit of submitted work, and help decide what should be funded or published. Peer review forms the central foundation of formal and implicit quality assurance in science, and have become an institutionalised way of evaluating what research to publish and fund (Schneider et al., 2021, p. 5). As publications and grants have increasingly become a type of "currency" for scientists, systems of peer review have also become an important mechanism for reward and recognition. Another institution of recognition are the norms of citation practices. How we cite other researchers, use the extant literature, and implicitly award credit to those we reference is also a core channel for distributing rewards in the scientific system (Schneider et al., 2021, p. 6).

However, the practice of peer review and citing behaviour is not only a neutral search for best or most relevant research. Recognition processes are also susceptible to cumulative advantage or "Matthew effects". First formulated by Robert K. Merton, "The Matthew Effect", deriving its name from the St. Matthew gospel, posits that rewards in science are primarily distributed through the recognition from peers. Ideally, recognition accrues to those who embody the norms of research by making genuine original contributions to science (Merton, 1957, p. 639). Merton's primary insight was that although major contributions to scientific advancement are recognised, peer-recognition is often distributed with a significant gradient (S. Cole & Cole, 1967, p. 382; Merton, 1968, p. 56). Well-established scholars tend to accumulate recognition through publications, awards, and research grants to a higher degree than less well-known researchers, irrespective of their talents or the intrinsic merit of their contributions (Bol et al., 2018, p. 1; Goldstone, 1979, p. 385). Merton later generalised the Matthew Effect as an example of cumulative advantage, specifying a process of increasing inequality and stratification, with (i) highly right-skewed distributions of awards, recognition or resources and an (ii) increasing dispersion of these outcomes over time.

When resources are allocated with a strong gradient, it gives rise to skewed distributions of publications, citations, prizes, and most importantly research funding. This is the first part (i) as described above and often argued to be one defining implication of cumulative advantage in the science system (DiPrete & Eirich, 2006, p. 274). The second part (ii) of cumulative advantage processes, predicts that highly skewed distributions of resources will reproduce over time and create lasting inequality. Thus, cumulative advantage implies that (adapted from DiPrete & Eirich, 2006, p. 280):

1. The growth rate of funding depends on current or past levels of funding.

2. Small differences in funding levels at an early stage grow larger over time.

3. The general inequality in funding should grow over time as funding differences accumulate

The process is considered to be self-reinforcing as recognition signals competence to those who distribute scientific resources like grants (Reskin, 1977, p. 493) through peer-review. Cumulative advantage works through two feedback-loops. Firstly, recognition motivates researchers to seek further recognition through publication and grant application, which leads to increased access to money, time, positions, and competent research assistance (Allison & Stewart, 1974, p. 597). Secondly, it leads the less well-off to become more disadvantaged. In other words, researchers rich on recognition and resources accumulate riches at a rate that makes the poor relatively poorer (Merton, 1968, p. 62).

To Merton (1968, pp. 57–58), cumulative advantage has, on average, a detrimental effect on the individual research career, as relatively unknown scientist are afforded less recognition than well-known ones, even for comparable achievements. But from the perspective of the whole scientific system, the unequal accruement of recognition may be entirely functional. It serves to heighten the visibility of important work and new discoveries (Merton, 1968, p. 59). Cumulative advantage in recognition, both through peer review of articles, grants and prizes or the accumulation of citations, serves to create a "productive" path dependency in the supply of science.

In the context of research funding, Mertons insights are important for understanding how funding is distributed to individual researchers. Unequal recognition and strong differences in accumulation of later resources apply to grant competitions where peer-review functions as the main allocative mechanism. Researchers submitting funded and unfunded proposals of almost equal perceived quality differ in their subsequent propensity to obtain funding and tenured positions (Bol et al., 2018). Strong differences between funded researchers also exists, as those with many grants tend to remain in the top of funding distributions, unlike researchers with only one grant (Katz & Matter, 2020).

The unequal attribution of recognition is likely to bias grant competitions against researchers with less "eminent" records in terms of publication outputs or citations. This may carry over to what type of research is funded, if highly rewarded scientists also tend to study a narrow selection of topics or are situated within a few privileged specialities and disciplines. Moreover, peer review is never conducted from a position of full information. Peer reviewers will be limited by their insights into the different subjects addressed by applicants, and may instead resort to citation metrics for discriminating between applications. Researching certain topics can grant applicants a clear advantage in meeting these expectations and increase their likelihood of having their projects funded (Stoeger et al., 2018). The functioning of peer review may also affect the selection into grant competitions. Researchers with less extensive records of accomplishment may self-select out of grant application writing because they anticipate rejections. Self-selection out of grant competitions may skew which topics are present in applications or what specialities and discipline are represented. Finally, initial differences in resources may induce positive feedback for the well-off researchers and negative feedback for those with less initial funding. Already having project funding could free up time for further grant applications and provide results or new questions for more projects.

Concentration at the individual level may then carry over to the level of research content. Researchers in the top of the funding distribution are likely to study the most "fundable" topics, while the less well-off have much fewer resources to investigate other topics. What research is conducted could then become a function of what a narrow funding elite considers worthwhile. How this plays out in practice is, however, not straightforward and is ultimately an empirical question.

From the perspective of cumulative advantage, concentration of funding on certain topics, specialities and disciplines is a function of concentration on individual researchers brought on by differential allocation of credit and recognition (e.g. through peer-review). These patterns of concentration can have both negative and positive consequences. The negative consequences may include a lack of diversity in who and what is funded, leading to a lack of broader perspectives, methods, or approaches. We may worry that strong concentration can diverge funding to "popular" areas of science even if other areas possess equal opportunity for discoveries. The positive consequences could be a focus on the most important topics or disciplines, with the best chances of furthering scientific advancement. This notion crucially depends on how science advances most effectively. Are important discoveries mostly made by a small elite of eminent scientists or by a broad span of the researcher population (J. R. Cole & Cole, 1972; M. H. MacRoberts & MacRoberts, 1987; Turner & Chubin, 1976)? This question notwithstanding, cumulative advantage surely is an important driver of what is funded, and highlights that we should not only worry about the degree of funding concentration, but also what type of research content funding is concentrated in.

3.4 External influence and prioritisation

Despite the important role of internal processes of reward and recognition, many external stakeholders (politicians, funders, business interests, etc.) and researchers themselves question the self-organisation hypothesis. In their view, a distribution of funding relying only on self-regulation and internal peer-review is unlikely to lead to societally optimal allocations of resources. Instead, some level of external prioritisation is perceived to be necessary to guide research funding allocations towards the right composition of research with wider societal applications. Polanyi's contemporaries argued that the distribution of resources to different disciplines and topics could not be understood by merely resorting to the scientific merit of that area. According to these arguments, we must also take its technological merit and social merit into account (Toulmin, 1963; Weinberg, 1963). For example, Alvin Weinberg, director of the Oak Ridge National Laboratory, argued that

> ”Science must seek its support from society on grounds other than that the science is carried out competently and that it is ready for exploitation; scientists cannot expect society to support science because scientists find it an enchanting diversion. Thus, in seeking justification for the support of science, we are led inevitably to consider external criteria for the validity of science–criteria external to science, or to a given field of science.” (Weinberg, 1963, p. 163)

Weinberg's divergence from Polyani's view of the scientific supply and demand functions is thus the explicit acknowledgement of a role for societal needs and applications. From this perspective, applied topics and disciplines are invested in because of their impact on societal problem solving, while more basic sciences and topics function as an overhead cost on this endeavour (Hellström & Jacob, 2012).

Furthermore, it should be stressed that Polanyi proposed the idea of funding distributions and topic choice as self-regulating processes in a vastly different institutional context, where competition for research funding was less pronounced and widespread than today. Prioritisation of funding by governments have changed drastically since then, and new private funders have begun investing heavily in basic research as well.

Likewise, the move towards large-scaled, earmarked, public funding to address "grand societal challenges" (Cassi et al., 2017; Wallace & Ràfols, 2015, p. 90) have become a prominent science policy tool for aligning science supply with societal needs. In addition, private research funders in the pharmaceutical industry have transformed their practices from a focus on in-house applied research to stronger external cooperation with university researchers (Ràfols et al., 2014). The mechanisms of self-regulation also seem unable to always reconcile supply and demand. Even if the scientific system is responsive to societal problems, clear misalignments exist. For example, investments in disease research is often misaligned with the health burden and death toll of individual diseases (Ràfols & Yegros-Yegros, 2018; Yegros-Yegros et al., 2018; Yegros-Yegros et al., 2020).

Therefore, we also need to recognise the external influence of policy makers and funders of science in the shaping the supply and demand functions, and ultimately the allocation of funding. While studies show that researchers retain strong autonomy over their chosen discipline and the topics they pursue, they are also constrained in their choices by external expectations. External influence on what type of research is conducted is generally perceived to have intensified as scientists have become more reliant on extramural project funding instead of recurring block funding for universities. This development has given research funding organisations, both public research councils, charities, and private foundations, more authority over research goals and content (Luukkonen & Thomas, 2016, p. 100; Whitley et al., 2018, p. 110).

Researchers' increased reliance on competitively awarded funding makes them susceptible to changing incentive structures exactly because so much of their work depends on obtaining it. Opportunities for funding are opportunities for continuing exciting work and contribute to furthering science. For this reason, actors external to the scientific process itself are likely to exert growing influence over the topics, specialities, and disciplines being furthered, as they may alter incentives for studying particular things. These dynamics of course vary from field to field, depending on how necessary competitive funding is for a given line of research. Studying topics within animal science can be highly dependent on funding from private companies and access to farmers, while political philosophy may be relatively more insulated.

Past research have highlighted the *politica*l influence on what is funded through strategic agenda setting by lobbying for certain funding allocations (Best, 2012a; Hegde, 2009; Hegde & Mowery, 2008; Hegde & Sampat, 2015; Reardon, 2014). Importantly, it has shown how political influence can translate into what type of science is conducted and what remains "undone" (Hess, 2009; Hess, 2015; Frickel et al., 2010, p. 4650). However, here I restrict my attention to funders' use of own resources. This view is broader and encompasses both private and public funding organisations. Hence, it translates into a focus on how both the state and private funders actively shape research content through competitive funding competitions targeted at strategic scientific areas of interest. How such funding arrangements systematically influences what is studied and what is not, is also less well understood (Gläser & Laudel, 2016).

To reiterate, many public research councils or government agencies are using funding programmes to influence the content of the projects they fund. They target funding towards "strategic" topics or research areas and elicit project proposals for specific problems, which are linked to governmental policy priorities (Gläser & Velarde, 2018, p. 2; Whitley et al., 2018, pp. 110–111). Funding programmes with thematic priorities have accordingly become a wide-spread policy tool to prioritise scarce resources for a growing scientific system (Gläser & Laudel, 2016, p. 126). The funding landscape has also diversified with a steady stream of new research funding organisations in many countries. Charities and private research foundations have become important sources of funding for many researchers, and these funders have their own priorities for research content (Gläser & Velarde, 2018, p. 2). Common to many contemporary narratives, is the notion that external interests almost completely determines the distribution of research projects across topics and specialities.

In the following, I discuss how increased thematic targeting of research funds by public and private funding organisations may impact problem choice and the wider distribution of funding to topics and disciplines.

3.4.1 Thematic priorities

The introduction of thematic priorities and targeted funding schemes can be expected to put in motion two paths for changing what topics researchers engage in and the wider distribution of funding to different topics. The first path are the *direct* effects of earmarking funding. When researchers struggle to finance their different lines of research, targeted programmes incentivise individuals to change their topic focus to better match these programmes. Scientists may discontinue some research lines where little funding is available, while strategically proposing projects within prioritised areas (Laudel, 2006b, p. 496; Laudel, 2006a, p. 390; Leisyte et al., 2010, p. 275; Luukkonen & Thomas, 2016, p. 123). A lack of less targeted funding has left some researchers feeling forced to meet strategic funding calls instead of applying for research they otherwise had in mind (Nuffield Council on Bioethics, 2014, p. 23).

The second path is shaped by the *indirect* effects of targeted programmes. Thematic priorities can have an indirect impact on what researchers, who apply to targeted programmes, work on. Specific funder priorities may shift the general incentives and collective priorities in the funding system. This can be the case when researchers work on projects funded through targeted mechanisms but also other projects. The thematic priorities from one project may spill over to future projects because a problem choice "lock in" research work in the foreseeable future (Leisyte et al., 2010, p. 281). Similarly, when a lead researcher attains funding, it also tends to affect the topic choices of their Ph.D.-students and postdoctoral researchers who may also shift focus to areas that are more fundable (Leisyte et al., 2010, p. 276).

Another indirect effect of targeted funding is how it could signal what is considered fundable. Targeted funding can signal to the scientific community what areas and topics are important and where to expect future possibilities for funding. Targeted priorities are especially argued to solidify emerging fields such as nanotechnology or synthetic biology (Gläser, 2019), but then continuing to skew research efforts through continued investment in few strategic areas (Nature, 2003).

It is however important to notice that earmarked or targeted funding schemes, where public funders use topic, speciality, or disciplinary priorities as part of the grant proposal review, can have differing effects on the concentration of funding. Potentially, it could also counteract concentration by allocating funding to less funded areas or encourage researchers to work across disciplines or take up new topics. Conversely, if priorities strongly align with priorities within the scientific system it may potentially exacerbate existing patterns.

3.4.2 Private funder priorities

In conjunction with the increased use of competitive funding instruments by governments, the system of scientific funding has also seen a large diversification of funding organisations. While private companies have long sponsored research within universities, these collaborations have often been contract-based and not open competitions as in research councils. However, this has increasingly changed as the funding landscape has become more diverse and heterogeneous, with research depending on, but also made possible through, many different extramural sources, such as charities, private research foundations, and trusts (Heinze, 2008; Whitley et al., 2018).

The diversification of funders, both public and private, could be expected to ensure a more balanced distribution of funds, because each funding organisation may differ in their priorities (Dasgupta & David, 1994; Nelson, 1959). This functional division of funding tasks should supposedly offer more diverse opportunities for topic choice and funding acquisition, and secure resources for more disciplines.

However, this is not an inevitable outcome. Just as public research councils prioritise their funding, new private-sector opportunities come with strings attached. It is equally possible that a diverse funding environment increase the dependence of researchers on particular funding organisations, or these organisations' expectations regarding content limit the accessibility of funding sources (Gläser & Velarde, 2018, p. 2). Also, priorities by these funders may have *indirect* consequences for the funding system as a whole by shifting incentives in whole system (see e.g. Best, 2012a, on direct vs. indirect effects of private priorities). Having access to funding from both private and public funders could lead researchers to focus their work on topics prioritised by private funders, and then apply for public funding with similar types of research problems.

Private funding may also contribute more *direct* effects on funding concentration. Grants from private funders may contribute to a self-reinforcing fortuitous funding circle for individual researchers. Grants from private sources may set off an accumulation of funding for successful applicants and the priorities within these grants may guide subsequent projects funded by public research councils.

3.5 The interaction of internal and external drivers of concentration

The previous sections of this chapter have discussed different perceptions of how funding becomes concentrated within a few disciplines, specialities, and research topics. Each of these perceptions have a different view on how the supply of sciences is affected. Is it mainly an internally driven process of self-organisation and may the internal institutions of peer review be prone to rewarding certain researchers and certain research content with an unproportionally large piece of the pie? And how does the expectations of external research funders influence these internal processes and shape what is funded?

In answering these questions, we should heed two qualifications. Firstly, each of the three perspectives point towards important mechanisms. Researchers are in most cases free to select topics and many are attentive to the development in their specialities and disciplines. They are, however, not blind to the inner workings of grant competitions and functioning of peer review, and often know how to tailor their research to what funders expect. Internal allocations of recognition are furthermore fundamentally tied to external expectations and vice versa. Many private research funders use peer-review in their appraisal of funding applications, and attainment of grants from private funders are often perceived as particular prestigious in the eyes of researchers (Hallonsten & Hugander, 2014). Particular researchers may also participate in setting priorities for both private and public funders, and are often members of boards or peer-review panels responsible for allocating funds. Finally, external funders are equally interested in funding researches perceived as excellent. Hence, internal and external mechanisms may interact in various ways.

In figure 3.1, I provide a very stylised representation of the three perspectives and the ways in which their proposed mechanisms could potentially influence each other. [1]

[1] This synthesis model builds on arguments presented in K. Gross and Bergstrom (2021, pp. 1–2).

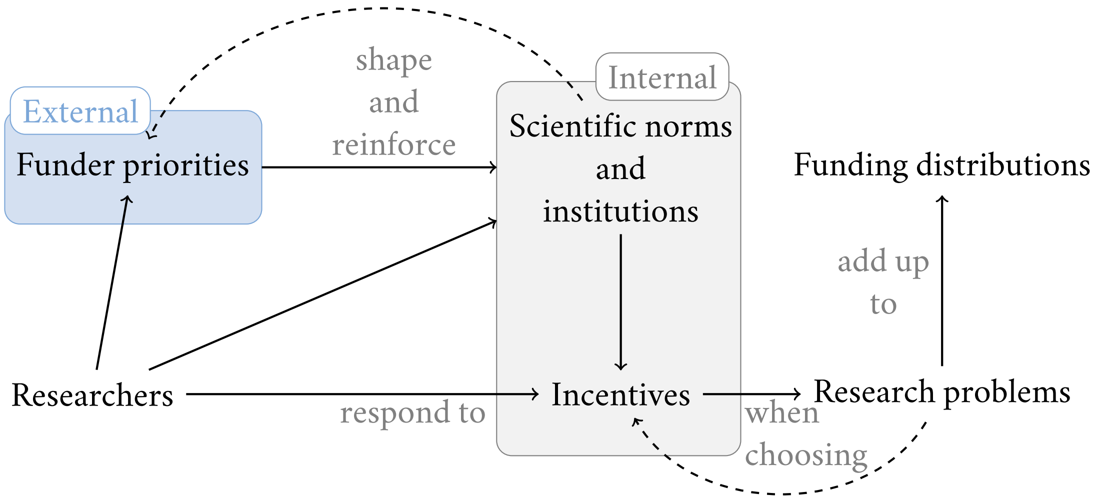


Figure 3.1: Theoretical model. The interaction of internal and external drivers of scientific supply and demand in funding.

The model highlights that no single explanatory frame can adequately account for observed patterns of concentration and dispersal of funding towards topics. The question of how the proposed internal and external mechanisms interact within different areas of science and in different national contexts to create more concentration or dispersal is largely an empirical matter. However, the model can help us derive some theoretical proposition on how the different parts interplay in ways that may either strengthen or weaken the degree of concentration at the research content level. In short, internal and external factors can enhance their individual effect of funding concentration thereby creating strong positive feedback effects, or can work counter to each other and result in stasis or negative feedback.

3.5.1 Positive and negative feedback mechanisms in funding

Focusing on internal norms and external priorities, we can identify several positive feedback mechanisms that may lead to increased concentration. Strong competition for responsive mode or non-prioritised grants may increase topical concentration when similar quality criteria are used in peer review processes. A lopsided focus on past performance through funding, publication, and citations is accordingly likely to increase incentives for studying particular topics and specialities, where one can already demonstrate competence. If many funders apply similar criteria, and peer-review/funding decisions are made independently of each other, the same researchers may prevail across competitions, which may in turn reinforce concentration. Under such conditions, trends toward concentration is inbuilt within the system, and arises due to internal factors.

A more external driver of concentration could be the formulation of similar priorities across many external funders. If funders tend to prioritise similar disciplines and topics it increases the incentives for studying specialities and topics falling within these priorities. The well-funded and successful researchers are likely to hoard funding in order to avoid having to change topics (Gläser, 2019). Concentration then becomes a self-reinforcing process. As Gläser puts it: "This buffering by applying for more grants becomes more important under conditions of decreasing success rates of competitive grant funding, but at the same time contributes to that decrease" (Gläser, 2019). Moreover, it may allocate a large share of the total amount of competitive research funding to a specific type of content.

But such feedback mechanisms may also work in conjunction. Funders may prioritise topics or disciplines, which are already highly funded in responsive mode grant competitions. Allocation patterns created on a backdrop of cumulative advantage signal to funders what type of research is worth investing in and where strong opportunities for discovery exists. Unequal reward patterns in non-prioritised grant competitions may give feedback to the priorities set by external funders, reinforcing positive feedback for researchers working within highly funded areas.

External priorities can also become dominant within the scientific community itself. Large investments by private funders may create pockets of strong research performance in some particular areas of science. When large investments in particular topical areas lead to larger resources and capacity, it becomes easier to attract funding from other sources. A well-funded research milieu may signal competence to other funders, create a strong basis for related projects with similar focus, and result in a strong track record for the scientists involved. Research centres established with an initial large investment often attract substantial additional funding from other sources (Bloch et al., 2016), and often result in strong increases in publications and citations within a research area (Ida & Fukuzawa, 2013; Rogers et al., 2012). Similar cycles of capacity feedback can also take place when external funders increase grant sizes. This can result in less topical diversity because more money is focused in one large project rather than many smaller ones, and may come at the expense of other areas (Bloch & Sørensen, 2015, p. 36).

However, internal norms of recognition and external priorities could also credibly lead to less concentration and more dispersal of funding distributions through processes of negative feedback. Following the logic of self-organisation, researchers themselves could gravitate towards other areas with less focus in order to facilitate new discoveries and claim priority for these (Merton, 1957). Funders may also periodically change portfolio strategy in response to new fields emerging. Some level of moderate dispersal will also remain stable due to the politically set distribution of resources following teaching obligations and Ph.D.-education.

Similarly, funders may at some point counteract strong concentration effects of the scientific reward system by earmarking funds for neglected topics or topics raised by interest organisations or other stakeholders (Best, 2012a; Best, 2012b). This has often been the case with strategic funding councils (Braun, 1998, p. 818). Even with thematically targeted grants, the scientific community often have a large leeway in influencing the outcome. Strategic priorities set by funders may be heavily influenced or even directly raised by researchers themselves, who are closest to identifying neglected opportunity (Braun, 1998, pp. 818–819). Moreover, individual researchers may be able to circumvent funder priorities, when these are vague or broadly defined, by repackaging and "window dressing" their project proposals to only appear as matching a given priority (Gläser, 2019; Laudel, 2006b; Leisyte et al., 2010; Luukkonen & Thomas, 2016). Changes in the topics that researchers address can also be temporary, or perhaps applicants are already engaged in prioritised topics and disciplines (Myers, 2020). A final possibility is that researchers simply diversify their portfolio of projects, and follow both new, prioritised, topics and a core set of topics thereby not changing overall distributions much. These latter points may admittedly also draw funding towards already well-funded topics and disciplines thereby potentially increasing concentration.

3.5.2 Some unanswered questions

Based on the interaction between external priorities, scientific norms or institutions pertaining to how recognition and rewards are distributed, and the self-regulating community of researchers, we can potentially imagine different empirical patterns for the distribution of funding over research content. When positive feedback mechanisms dominate, it will lead to increasing concentration over time, and a situation reminiscing cumulative advantage, where some topics or specialities are increasingly funded, while other are relatively less so. When both positive and negative feedback are at work, we are likely to observe less diverging funding paths. Instead, positive feedback may lead to initial concentration, which will then remain relatively static over time. This also results in strong funding differentials, but not to an increasing degree as cumulative advantage would predict. Lastly, negative feedback may lead to more dispersal when new funder priorities are enacted through government budgets or investments, or when sheer change in scientific opportunity shifts researcher priorities and incentives. This leaves some unanswered questions guiding the empirical parts of this dissertation.

Firstly, to assess the predictions above, we need to know: *How is competitive research funding distributed across individual researchers and categories of research content?* Not only does this question imply an empirical investigation of the distributional characteristics in scientific funding, but also what topics, specialities or disciplines funding are concentrated in. The question addresses a baseline descriptive background on which the whole dissertation is built, and is examined from different angles in papers A, B, and D. Where paper A and D primarily address the question of how skewed distributions are, paper B further examines which disciplines and specialities (i.e., disease areas) are primary beneficiaries in funding competitions.

Secondly, to assess the degree of path dependency in funding allocations, I qualify the above by asking: *How static are funding distributions, and how do individual and content distributions interact?* Paper A focuses on the first part of this question by investigating how topics and specialities situated in different parts of the funding distributions fare over time, and whether past funding levels are predicative of later funding levels. Paper D instead takes on the latter part in showing how the strong concentration of funding at both topical, disciplinary, and individual level appear interconnected.

To assess the influence of external priorities, two questions are especially relevant. The third question is: *Does targeted government funding shift individual and collective choices for topics?* In paper C, we attempt to discern the role of targeted funding in shifting what researchers choose to study, by analysing the effect of earmarked grants on individual researchers' topic portfolios. Paper D instead grapples with the wider impact on the funding system by showing how targeted funding often aligns with distributions from responsive mode grants.

The final question hones in on the role of private funders: *Does private funding shift the overall distribution of funding to topics, specialities, and disciplines?* In paper B, we investigate how funding is allocated across disease areas and compare the patterns to what one would expect given each disease's societal health burden and how it lines up with interest in the pharmaceutical industry. Paper D instead compares funding patterns by a range of private and public research funders to assess the degree of overlap between the priorities of governments and private interests.

4. Studying funding concentration: Data and approach

This chapter describes two of the overarching methodological challenges ingrained in the study of how funding is allocated. Firstly, commonly used data sources are often ill-fitted for answering even descriptive questions pertaining to the type of research funded via competitive funding schemes. Official funding statistics lack detail in terms of the disciplinary and topical categories and do not allow for comparisons between different funding organisations. Conversely, detailed data collected by researchers often lack sufficient coverage, and do not allow for analyses of the broader research funding system.

Secondly, while some data sources identify the institutions or researchers to which competitive funding is awarded, very few include information on the type of research that was funded. I attempt to overcome these challenges by using information on approximately 120,000 individual grants awarded by 22 research funding organisations in Denmark and the United Kingdom. I connect these grants to publications produced by grantees in order to assess the range of topics, specialities, and disciplines that the grants cover. I start out by discussing the competitive funding systems in Denmark and the United Kingdom, the shortcomings of existing data sources for funding statistics, and detail the data collection that forms the basis of this dissertation. Finally, I discuss how I categorise research content from publication data and relate these efforts to the discussion of topical structures from chapter 2.

4.1 Competitive research funding in Denmark and the United Kingdom

Throughout this dissertation, I rely on grant data from Denmark and the United Kingdom. Both countries have a common outset in what is known as the dual-support system of research funding. General institutional "floor" funding is allocated to institutions and combined with competitive funding awarded to individual researchers or institutions through grants and contracts from research councils, governmental departments, and private funders (Aagaard, 2017; Geuna & Martin, 2003). In the UK, this system has undergone fundamental changes over the years (Geuna, 2001). The first public research council (the Medical Research Council) was founded before the Second World War in an effort to separate governmental research contracts on specific scientific problems from projects of a more general concern. Throughout the 1960's, similar research councils were founded for other fields culminating with the Arts and Humanities Research Board (later the Arts and Humanities Research Council) in 1998 (Agar et al., 2019). Similar developments took place in Denmark, but at a slower pace. Throughout the post-war years, university block grants were the dominating source of research funding. Competitive research funding schemes were slowly introduced in the wake of the Second World War with the establishment of the Technical-Scientific Research Council in 1946 and the State Science Foundation[1] in 1952 (Aagaard, 2011, pp. 169–170). Political demands for more selective allocations of resources culminated in the establishment and institutionalisation of the research council system in 1968 (Aagaard, 2017, p. 285). Later developments in both countries have led to a higher degree of competition for funding. The floor-funding models of both countries have been revised to incorporate some evaluative or performance-related element, either through the UK Research Assessment Exercises/Research Excellence Framework or the Danish use of the bibliometric research indicator or the pegging of block grants to the number students, Ph.D.-students, and an external grants indicator (Geuna & Martin, 2003, pp. 280–281; Stage & Aagaard, 2020, pp. 852–853).

However, both funding systems have also witnessed strong displacement dynamics between the two funding streams in terms of balance and composition. In Denmark, the percentage of funding from external sources increased drastically between 1970 and 1990, from 11 % to 35 % (Aagaard, 2017, p. 298), and currently this share has grown to an estimated 45 % (The Danish Council for Research and Innovation Policy, 2020a, p. 20). The research councils are also taking up an increasing share of the funding base for UK universities. While both institutional funding and research councils funding have grown between the 1990's and 2000's, the share of funding being allocated in open competition through research council application calls have almost doubled (HM Treasury, 2003, pp. 80–82). Besides the increasing reliance on competitive funding, the two countries have seen varied changes in the composition of this funding.

Firstly, governments have introduced more strategic or thematically targeted funding. In Denmark, thematic funding has largely been implemented through two processes. Between 2002 and 2005, the Danish government established three new strategic research funding channels alongside the traditional research councils: The Council for Technology and Innovation, the Strategic Research Council, and an Advanced Techno-

[1] "Statens Almindelige Videnskabsfond"

logy Foundation. These have later been merged to form the Innovation Fund Denmark (established 2014), and in 2019 the Independent Research Fund was also tasked with distributing thematic funding.

In the UK, strategic research funding stands in opposition to responsive mode funding (or researcher-initiated funding), and the former was typically the domain of government departments. However, the merger of research councils into the UK Research and Innovation (UKRI), and the relative decline of strategic funding to departmental contracts and in-house research, have led the research councils to become a swiss army knife of UK science policy (Agar et al., 2019, pp. 12–13). Strategic priorities for certain research themes have been implemented through e.g. targeted mode funding (as opposed to responsive mode) in the Engineering and Physical Sciences Research Council (Kearnes & Wienroth, 2011), the responsive mode priorities implemented in the Biotechnology and Biological Sciences Research Council (BBSRC), or after the formation of the UK Research and Innovation (UKRI) through council-wide themes and the Strategic Priorities Fund.

Secondly, both countries have seen a rise in external funding from private entities, such as research foundations and charities. The UK funding landscape includes some of the world's wealthiest foundations involved in sponsoring research. While government funding has declined somewhat, universities have diversified their funding sources by taking in more funding from charities and industry (Kundu & Matthews, 2019, p. 612). Denmark have experienced similar developments. Recently, private research foundations such as the Novo Nordisk Foundation, the Lundbeck Foundation, the Carlsberg Foundation, and others have massively increased the amount of competitive funding they provide for research, which has introduced a large amount of funding tied to corporate interests and priorities into the Danish system. Recent analyses from the Danish Council for Research and Innovation Policy suggest that around 50 % of Danish research funding are competitive funds, and 20 % stem from private research funding organisations (The Danish Council for Research and Innovation Policy, 2020b). This development is also likely to continue. Around two thirds of competitive research funding in 2017 and 2018 were from twelve of the largest private research foundations, and the remaining third came from public research councils (The Danish Council for Research and Innovation Policy, 2020a, p. 17).

In the analyses throughout this dissertation, I use data from both countries, but in different capacities. I do not conduct a strict comparative study of the entire competitive funding systems in both countries. Instead, I leverage information on grants from both cases in different ways. The Danish data covers all the primary external funders, including private and public organisations, as discussed below. I use these data to investigate the broader distributions across an entire funding system and to compare what is funded between public and private sources. Data from the UK instead cover only the major public sources of funding: The UK Research Councils. These data instead provide more detailed information about individual grants and their outcomes, and are suited for investigating the finer details of individual researchers and their responses to having received a grant.

4.2 Studying research funding distributions

The question of how research funding is allocated across topics has drawn attention in several parts of the science and higher education policy community. More recently, such policy debates have centred around discussions about the type of "funding portfolio" funders should prioritise (e.g. Waltman et al., 2019). For example, funders have set up funding schemes for addressing grand societal challenges such as climate change, antimicrobial resistance, global pandemics, and food security to better align science supply with the needs and demands of the wider society (Wallace & Ràfols, 2015). However, to judge the success of such programmes or the wider allocation of resources in general, there is an increasing need for detailed data on who funds what. Such tools are however seldom available from traditional data providers. Consequently, the available knowledge on the character and effects of competitively awarded research funding remains sparse (Aagaard et al., 2019).

4.2.1 Problems with existing data

Different national statistics offices, international organisations, and researchers have provided information on the size and direction of research funding (see a selection of funding data sources in table 4.1). These accounts however suffer from at least three limitations.

Firstly, they often lack the necessary detail on how different areas of research are prioritised by funders. Often, funding statistics are also not able to differentiate between the different providers of external research funding. The OECDs Main Science and Technology Indicators list the sources of external funding[2] at the sectoral level, distinguishing only between the government sector, higher education sector, the business sector, and the non-profit sector. A little more detail is included in data from e.g. Statistics Denmark, Danish Universities, the Higher Education Statistics Agency (HESA), and the UK Office of National Statistics (ONS). Here, the sectors are subdivided and show funding amounts from government agencies, municipalities, and public research coun-

[2]Note that this includes both funding awarded through contracts as well as competitive funding.

cils, while private funding is often only separated into either the business sector or the private/non-profit sector.

The same data providers seldom provide information on who receives research funding. The most detailed accounts only provide information on how much funding is allocated to research-performing organisations such as universities, public research laboratories, and governmental research organisations. However, the most problematic aspect of these types of data is the missing information on the type of research conducted. Some organisations (e.g. HESA) provide no information of how funding is allocated across fields of science. The OECD provides funding expenditures by six fields and 39 subject categories. In a Danish context, Statistics Denmark provide 56 subject categories, but these categories differ immensely in terms of their level of aggregation. The medical sciences are represented by seven categories: Basic medicine, pharmacy/pharmacology, clinical medicine, odontology, care, social medicine and public health, and medical biotechnology. Sixty-eight percent of expenses are distributed to basic and clinical medicine, but this provides no indication on how much funding each medical speciality (e.g. cardiology, endocrinology, etc.) receives. In sum, we lack a well-defined and detailed granularity of topical and field-related allocation patterns.

Secondly, many researchers have collected their own sets of data concerning the distribution of funding. These are often much more detailed in terms of who awards money to whom, but lack broader coverage. A number of papers have investigated the concentration of funding on individuals or institutions within one research funding organisation such as e.g. the U.S. National Institutes of Health (Berg, 2012; Bowen & Casadevall, 2015; Hegde, 2009; Hegde & Mowery, 2008; Hegde & Sampat, 2015; Katz & Matter, 2020; Lorsch, 2015; Manton et al., 2009; Peifer, 2017; Stoeger et al., 2018; Wahls, 2016, 2018a, 2018b), the National Science Foundation ((S. Cole et al., 1981; Klavans & Boyack, 2017a), the UK Engineering and Physical Sciences Research Council (Ma et al., 2015), the Australian Research Council (Bromham et al., 2016), the Veni, Vidi, Vici Program in the Netherlands (Bol et al., 2018), and the Natural Sciences and Engineering Research Council in Canada (Fortin & Currie, 2013). Other studies include data on multiple research granting institutions in e.g. Quebec (Larivière et al., 2010; Mongeon et al., 2016). Nevertheless, the scope of these studies are limited to specific subsets of funding organisations.

Thirdly, many of the existing data sources and studies of funding allocations base their analyses on a purely institutional definition of research fields, topics and disciplines. This is to be expected as funding organisations decide on whether to fund research groups, departments, laboratories and centres, as opposed to entities purely defined in terms of cognitive boundaries (Martin & Irvine, 1985, p. 559). A pre-occupation with the institutional notion of topics has several drawbacks. For one, it tends to be a less precise

Table 4.1: Examples of funding data sources and their level of detail.

	Structure	Source	Topical category	Performer
OECD MSTI	R&D funding	Sector	39 subject categories	Sector
Statistics Denmark	External funding	Sub-sectoral	56 subject categories	Sector
Danish Universities	External funding	Sub-sectoral	None	Universities
ONS	External funding	Sub-sectoral	None	Universities
HESA	External funding	Public granting organisations	None	Universities
Collected data	External funding	Granting organisation	No principal limit	Grantee

way of assessing funding allocations. Statistics Denmark uses survey questions directed to around 700 research-performing organisations to gauge what disciplines receive what amounts of funding. These are most likely answered by administrative staff, and do not indicate what the funding was actually used for or the topical diversity within a research unit. Statistics provided by the UK research councils rely on researchers' self-report of what topics their grant addresses at the proposal state, and gloss over the later results. By relying on input-based categorisations of grants, these types of data may better reflect the type of research intended rather than the type of research actually conducted. While the former may be interesting in a study on how a researcher's topical focus shifts during a project (Boyack et al., 2018), it is the latter that speaks most to relevant science policy debates. In the following sections, I therefore use output-based categorisations based on publications by recipients of grants.

Another challenge related to topical categorisations based on who receives funding, is the question of how related different topics or disciplines are. When categorising the allocated amount of funding by looking at what institutional entities are the receivers, we know very little about the intellectual distance between these entities. As an example, we know that 12 % of research and development expenses in the Danish public sector took place within clinical medicine in 2018. At the same time, biology and biochemistry respectively made up 4 and 2 % of expenses (Statistics Denmark, 2020). Without knowing how similar these disciplines are, we will have a hard time investigating whether this is a sign of strong concentration. Perhaps biology relates very little to clinical medicine, and work with entirely different problems, which would signify greater diversity. Or perhaps the type of biological research conducted is closely related to medical research through studies of cell mechanisms and DNA, whereby the concentration is more pronounced than it appears.

To summarise, a closer investigation of the topical concentration of research funding demands more detailed data on who receives and distributes funding, what funding was actually used for, and how similar or different grants are.

4.2.2 Fine-grained data on grants in Denmark and the United Kingdom

The analyses presented in the following chapters rely on two unique databases of individual grants awarded to researchers in Denmark and the United Kingdom. Data on grants awarded by research funders in Denmark come from the in-house grant database compiled by the Danish Centre for Studies in Research and Research Policy (see Aagaard et al., 2019). The database covers 19,399 grants with a value of 52.9 billion DKK awarded to 7,539 grantees in the period from 2004 to 2016. The grants were awarded by 15 public and private research funders, including the Carlsberg Foundation, the Independent Research Fund Denmark (DFF), the Danish National Research Foundation (DNRF), the Council for Technology and Innovation, the Strategic Research Council, the Advanced Technology Foundation, the European Research Council (ERC), the Danish Cancer Society, the Lundbeck Foundation, the Nordea Foundation, the Novo Nordisk Foundation, TrygFonden, the Ministry for Higher Education and Research, the Velux Foundation, and the Villum Foundation. For the United Kingdom, I compiled a database of 97,861 grants awarded to more than 22,000 individual researchers between 1981 and 2017, with a combined value of 39.3 billion pounds. The British grants were awarded by the seven British research councils: Arts and Humanities Research Council (AHRC), Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC), Medical Research Council (MRC), Natural Environment Research Council (NERC), and Science and Technology Facilities Research Council (STFRC).

In both cases, grants from innovation-focused research councils, i.e. Innovate UK and the Innovation Fund Denmark, were left out because grants are often directed at firms, organisations, or larger innovation consortia and do not exclusively support individual principal investigators and research activities. All data were compiled through a combination of manual and automatic collection of metadata on individual grants, including the recipient, the institutional affiliation, the time of award, duration, value, title and abstract. For the Danish grants, individual funders were contacted and asked to provide information on allocated grants. Data from the British research councils where collected through automatic web-scraping of the RCUK Gateway to Research Database, the BBSRCs advanced grant search, and the EPSRC Grants on the Web.

Both grant databases include individual grants awarded to researchers with very different purposes. My main focus is to assess the distribution of research funding, and I excluded a number of grants with this in mind. In the Danish case, many funders award money to projects that are not research-related or aimed at producing a traditional research output, such as journal articles, books, and patents. The Danish data therefore excludes all grants aimed at training or education, as well as grants with a monetary value below 50,000 DKK. The majority of these small grants are small student grants for conducting a minor research project during a master's degree, or small travel grants for researchers attending conferences or workshops. A few grants were large centre grants awarded to a dean or a rector at a university. These were also excluded, because it was clear that these grantees were not the intended users of the grant.

For the British data, the full database initially consisted of 132,091 grants. I excluded grants from Innovate UK and grants from other research councils such as studentship grants, skill-development grants, partnership and communication grants (P&Cs), intramural grants, and grants not categorised. The remaining 97,861 grants were research grants, fellowships, infrastructure and equipment grants, grants to set up whole units or centres, third party grants, and training grants. Training grants are larger grants aimed at training Ph.D. students. This is arguably part of the research process, but in most cases, I also excluded these grants because the research outcome of funding these Ph.D. students are diffuse and vague.

All collected grants where assigned to a single principal investigator or grantee, even when explicit co-investigators were listed, and duplicate grants were removed. Focusing money on a single researcher will inevitably increase the level of funding concentration in the system as grants are almost never used exclusively by the principal investigator, but also by postdoctoral researchers, Ph.D. students and other temporary and more permanent staff affiliated with a funded project. Therefore, the data cannot account for how funding is spread across researchers after being awarded to a grantee. However, grants are awarded to researchers with the explicit purpose of acquiring resources to conduct a project, including hiring other researchers, or buying materials and equipment. Even if science is largely a team-based endeavour, it is often the higher ranking team members who receive the bulk of recognition and prestige from papers, prizes, and grants (S. Cole, 1970; Zuckerman, 1992).

In the two sets of data, each grantee was disambiguated to ensure that grantees with multiple grants across times could be linked. In the Danish data, researchers were manually disambiguated. In cases of common combinations of first and last names, a grantee was identified as unique by comparing their institutional affiliation and main research area. If this was not sufficient information to separate two similarly named individuals, these were treated as unique persons. In the British data, grantees were identified through their individual research council identification number, which is allotted during the application process. Furthermore, I used a combination of e-mail-addresses, institutional affiliation, and ORCID-IDs to resolve outstanding identification problems.

Information on the value of a grant was standardised to either Danish kroner or pound sterling, and recurrent grants with an annual pay-out was summarised and centred on the first year of the award. For the Danish data, grant amounts were kept at current prices. In the case of the British research councils, grant amounts were also

Table 4.2: Summary of grant datasets.

	No. of funders	Time period	No. of grants	No. of grantees	Grant value
Denmark	15	2004-2016	19,399	7,539	DKK 52.9bn
United Kingdom	7	1981-2017	97,861	>22,000	£39,4bn

standardised by adjusting for the consumer price index with 2015 as the base year. In spite of the large amounts of data, the two databases do not cover all competitive funding awarded in the time period, and cannot be treated as a complete picture of the competitive funding landscape in Denmark or the United Kingdom. It may not accurately depict the absolute amount of funding awarded to different individuals or topics, but it provides important insights into the relative prioritisation within the funding systems. Table 4.2 summarises the two datasets.

4.3 Categorising research content

Contrary to most systematic studies of research funding concentration, I primarily investigate how funding is allocated across topical and disciplinary groups and not individual research institutions or scientists. This ambition provides some strong methodological challenges. As argued in chapter 2, topics are not a set entity like a person, and we cannot argue that one conceptualisation of a topical structure is final, "true", or even the most appropriate. How topics are conceptualised and measured may have opaque consequences for the conclusions we draw from these structures. Do the sets and categories of grants I identify actually represent thematic groups? To what extent are these identifications determined by the properties of actual unobserved knowledge structures, and to what extent are they artefacts of the approach itself? These are questions, which I cannot answer, and which would be hard or impossible to answer altogether (Gläser et al., 2017, p. 983). However, two broad considerations are important to emphasise in order to understand how I conceptualise research content and categorise grants. First, the question of whether to use output or input information from grants. Second, whether to categorise grants from top-down categories, or instead create categories in a bottom-up fashion.

4.3.1 Output vs. input: Using publications to classify grants

A central choice in categorising research grants into topics and disciplines relates to the type of information about a grant to base this classification on. Even if the literature on how funding is distributed across topics is still sparse, some work on classifying grants have been undertaken. One way of classifying grants have been to use *input* information from the grant proposals, i.e. the content of a grant proposal. Klavans and Boyack (2017a) use text similarity measures to relate grants from the National Institutes of Health, the National Science Foundation, the CDC, NASA, and other federal research funders to 91,726 topics derived from 48.8 million research articles and reviews. The initial topics were created by clustering research articles by their direct citation relations, such that if a document A cites a document B, a direct citation relationship exists. A grant was then related to a topic by calculating the text similarity of a grant's title and abstract to the text in articles. Another option has been to use the references cited in a grant proposal (Boyack et al., 2018). Each article in the reference list then reveals something about the topical focus of the grant proposal, and can be linked to metadata from publication databases like Web of Science and Scopus. These metadata can then be used to create a topical or disciplinary classification, using either the inbuilt subject categories of these databases, or a tailored classification scheme via citation-relations, text similarities, etc. (Boyack et al., 2018, p. 454).

I instead used *output* information, i.e. the results of a grant, to classify each grant into topical and disciplinary categories. More specifically, I used either research publications explicitly linked to a grant, or publications produced by the grantee in the years following a grant's start date. Using output information to classify grants carries two advantages compared to using input information. First, input information is often classified and not available for a large selection of grants. Due to secrecy and data protection regulations, funding organisations are not willing or able to provide the full text and references of successful grant proposals. Output information, in the form of publications, are already in the public domain, and can be readily accessed. Second, input information reveal what topics a grant is meant to address, but not what the money was actually used for. Publications instead signify what the tangible outcome of a grant was, and what ended up being the main cognitive focus. To facilitate a categorisation of grants, publications were matched to grants in two ways.

4.3.1.1 Grant data from the United Kingdom

For the British data, I matched each grant to the output publications using two data sources. First, I used the RCUK Gateway to Research to identify self-reported outcomes of grants, and narrowed the search to only journal articles and reviews. Grantees report the outcome of their grants, including publications, and these are listed as linked to specific grants. The search yielded a gross sample of 305,247 publications directly linked to a specific grant. I then matched these self-reported publications to article-level information through the in-house version of Clarivate Analytics' Web of Science database hosted at the Centre for Science and Technology Studies at Leiden University. The Leiden database is an augmented version of the Web of Science with a thorough author-name disambiguation, which assigns individual identification numbers to publications from the same author. The identification numbers are algorithmically generated, and not a perfect representation of a single author. Some authors may have multiple identification numbers, and some numbers may cover more than one researcher. However, it represents a state-of-the-art approach to author disambiguation in a large-scale publication database that cannot reasonably be disambiguated manually.

This set of grant-linked publications yielded 180,100 documents, i.e. 60 % of the self-reported publication output. Therefore, I cannot cover the complete 97,861 grants with reliable information about their topical focus. Firstly, out of 97,861 grants, only 36,559 grants had self-reported publications as part of their outcome. This is the case because many grants awarded before 2006 did not report any publication output. Second, Web of Science has a much lower coverage of publications within the humanities and the social sciences (Mongeon & Paul-Hus, 2016). The Arts and Humanities Research council awarded 5872 out of 97,861 (6 %) grants, but I could only link 725 grants to publication output. Similarly, many grants within computer science and engineering report conference papers as output and is the primary mode of scholarly communication in some disciplines, which could not be matched to the Web of Science either.

4.3.1.2 Danish grant data

In the Danish case, no funding organisation have consistently classified grants into topics, and do not collect data on the output produced. Consequently, grants from Danish funders could not be linked directly to publications as in the UK data. Hence, to ascertain how funding is distributed across different types of research we relied on journal articles and reviews published by the grantees. To link each grantee to their publications, we use a combination of automated and manual matching, by comparing names, email addresses and institutional affiliations in the Leiden database version of the Web of Science. We matched 5,773 grantees representing 14,103 grants, to their publication profiles. Together, these grants account for around 80 % of the combined funding amounts attributed to individual principal investigators. Again, most of the non-matched grants were awarded to research within the humanities and social sciences.

To narrow the topical spread even further, I focused on publications 0–4 years after a grant was received. This approach leaves us with only 63.3 % (12,269) of all the collected grants, but should better reflect the topics a grantee was working on at the time of the award. Because these publication data do not represent direct output of a single grant, they are more uncertain than in the case of British research grants. It is reasonable to assume that they reflect the overall focus of a grantee, but likely underestimate the extent of funding concentration as a grantee's combined publication profile is likely broader than the output of a single grant (Aagaard et al., 2019). However, the direct link between grants and publications in the UK data is not without flaws either. Researchers themselves admit the use of some grant money on work that does not relate to the actual grant (see chapter 3). Moreover, grants may be used to fund non-specific work that covers a wide range of topics.

Lastly, I cannot cover all grants collected in the two datasets because publications related to each grant were not available. If this missing publication information were largely random, it would not be grounds for any concern. However, it is likely that some grants may not result in any publications because that was not the intent, or because a project failed to produce results worthy of publication. For this reason, the analyses put forth here will have a lower coverage of e.g. innovation-focused grants, which may aim to produce patents or refine already existing products instead of grants resulting publication output.

4.3.2 Top-down vs. bottom-up classifications: Measuring topics

The choice of whether to rely on input or output information from grants is a choice of what basis to use for classification and categorisation. The next choice relates to how we use this information to create topical and disciplinary categories. I argue that we can approach this task from either a *top-down* or a *bottom-up* perspective.

In chapter 2, I proposed that we could define topics in three distinct ways: Institutional, cognitive, and semantic. These three types of definitions closely fit the distinction between topics measured from the top and down vs. the bottom and up. The institutional notion of topics lends itself to a top-down approach to measuring topics, as set structures already institutionalised. One example of this is the American Physical Society's categorisation of research articles in physics: the Physics and Astronomy Classification Scheme (PACS). The categorisation is based on 10 broader disciplinary areas, encompassing nine sub-fields for each area, and a maximum of three more levels of granularity. For example, one top-level category is "Nuclear Physics" (20), with the sub-field "Radioactive decay and in-beam spectroscopy" (23), which can be further divided into e.g. "Electromagnetic transitions" (23.20) and finally into "Multipole matrix elements" (23.20.Js). The categorisation scheme was implemented in 1975 (Krumhansl & Trigg, 1975), and have seen both additions and changes, but the core parts of the categorisation rests on long term notions of how physics is subdivided. The categorisation is made and maintained by an organisation of physicists, with trained personnel developing the categorisation with input from researchers (Smith, 2019). In this way, PACS codes are top-down categories applied to pieces of research through institutionalised knowledge about the content of this research.

Cognitive and semantic notions of topics instead favour a bottom-up approach where data are not fitted into existing categories, but "learned" from the underlying

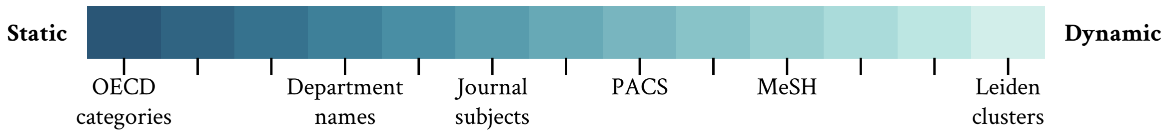


Figure 4.1: Static and dynamic topical categories as a sliding scale.

data structure. Here, the starting point is not the categories themselves, but the similarity and difference between data points, i.e. grant text or publications' reference structures. An example of this is how the Leiden Ranking, an annual ranking of over 1000 universities worldwide, uses algorithmically constructed research fields to characterise the topical profiles of these universities and to field-normalise impact indicators, such as citations (CWTS, 2020). The fields are constructed by laying out a network of publications by the universities, with each publication being a node and each node being linked by direct citations between them (i.e. the edges between nodes). Publications are then clustered into 4,013 topical clusters or communities through a modularity optimisation algorithm (Traag et al., 2019; Waltman & Van Eck, 2013; Waltman & van Eck, 2012). A complete overview of the procedure is beyond the scope of this dissertation, but a short description illustrates the steps taken. The algorithm clusters publications by optimising the modularity of the network, with the modularity being the difference between the actual number of direct citation links in a topic cluster and the expected number of such links (Traag et al., 2019, p. 1). Formally, the modularity can be described as:

$$\mathcal{H} = \frac{1}{2m} \sum_c \left(e_c - \gamma \frac{K_c^2}{2m} \right) \tag{4.1}$$

where e_c is the actual number of direct citation links in a topical cluster c. The expected number of links, if the links where randomly distributed, is given by $\frac{K_c^2}{2m}$, with K_c being the sum of the number of links to each node in the community c and m being the total number of nodes in the network. Finally, γ is a resolution parameter that guides the number of communities generated. Topical clusters are then created by optimising the function, using an algorithm. In the case of the Leiden Ranking, this is either the Smart Local Moving algorithm (Waltman & Van Eck, 2013) or the Leiden algorithm (Traag et al., 2019). The topical clusters are thus created "bottom-up" by classifying papers based on their direct citations of each other. However, the distinction between top-down and bottom-up, or institutional vs. cognitive/semantic topic conceptualisations are not necessarily clear cut. Instead we may think of this as a continuum of topic classifications that runs from the complete static top-down categories to completely dynamic bottom-up categories as shown in figure 4.1.

For example, the statistical categories employed by the OECD or Statistics Denmark are very static to ensure comparability over time, and are based on institutionalised notions of how science can be categorised. Institutional names, such as university departments, are less static and can change over time, but tend to do so infrequently. Topic categories based on scientific journals are even more dynamic as new journals are established and included in publication databases. Examples such as the APS' PACS codes show stability as the overarching disciplinary definitions seldom change, while lower level sub-fields are removed and added to the classification tree.

At the other end of the spectrum are the bottom-up, data-driven, categories used in e.g. the Leiden ranking. These categories can be constantly tailored by changing the resolution parameter to create more of fewer topic categories, the membership of different topics clusters change when new publications are included, and when citations between these change. Whereas statistical categories represent one way of dividing science into groups at a specific time, the data-driven groups represent the ever-changing nature of the scientific system.

Each of the endpoints in this continuum also have distinct advantages and disadvantages. The institutional conceptions of topics and disciplines provide clear and understandable labels for each category of research, and likely resonate with many scientists themselves. These are labels used by scientists themselves when describing their work or when interacting with other researchers. Labels such as "biochemistry" or "high energy physics" provides a well-known context and carries implicit meaning for people.

Bottom-up categories based of citation relations do not provide automatic labels, but instead offers a much more detailed and granulated view of the distribution of topics as it is at the moment. This high level of detail is also important when researchers and programme officers in funding organisations can differentiate between around 100,000 topics (Boyack et al., 2014; Klavans & Boyack, 2017a). Moreover, topics based on a citation network provides information on how closely two topics resemble each other by way of their relative positions in the network. In this dissertation, I try to leverage the advantages of both the top-down and bottom-up categories, while also combining some of their features.

4.3.2.1 Top-down categories describe what is funded and what is not

Throughout the analyses presented here, I use institutional conceptualisations and top-down classifications of disciplines to describe what sort of research receives a lot of funding and the type of research that is more marginally funded. In Article B, we use a combination of two top-down categorisation schemes.

First, we use the 39 statistical categories defined by the OECD to give an overall view of what disciplines are funded more or less across the 15 Danish funding organisations.

Contrary to how the OECD use these, we modify the categories in two ways. For one, we apply them to the individual articles produced by a grantee, and calculate the proportion of her publications devoted to each of the categories. In this way, we can represent each grant as a mix of different disciplinary content. Then we use the citation traffic between each of the 39 categories to relate their intellectual similarity. This is essentially a mix of both a top-down and a bottom-up approach to disciplinary categories. In doing so, we show that the biological sciences receives a large bulk of available funding in Denmark. It is important though to note that this skewness in the funding distribution is partially associated with the relatedness of biological sciences to both basic and clinical medicine. Researchers within biology cite these two disciplines more than e.g. chemistry.

In the second part of the article, we take a closer look at the biological and medical sciences by analysing the spread of funding across 143 disease areas. The disease areas are based on the United States National Library of Medicine's Medical Subject Headings (MeSH), which are used to tag articles in their PubMed and MEDLINE databases with information about diseases, organisms, or chemicals and drugs addressed in an article. The MeSH categorisation covers thousands of individual categories across its 16 top branches. In the disease branch, there are around 10,923 individual terms used to describe an article. We simplify these by translating MeSH terms into the 143 disease categories used by the WHO when calculating the burden of different diseases across the world.

Lastly, I use the Web of Science's 250 subject categories to describe the distribution of funding across disciplines in Article D. I set out by simply using the categories as is, i.e. the top-down definitions of what articles fit into what categories. I do however also refine this categorisation. The Web of Science allocates some of the world's most popular cross-disciplinary journals to one topic category: Multidisciplinary Science. This choice results in a categorisation where every paper in e.g. Nature, Science or Proceedings of the National Academy of Sciences are grouped into one category. Given that these are very prestigious and popular journal choices, the most funded category tend to be the multidisciplinary one, despite many papers being purely within one discipline, e.g. astrophysics or molecular biology. To mitigate this, I re-allocate all multidisciplinary papers based on the most commonly referenced subject category in their list of references (for a detailed approach see Milojević, 2020).

4.3.2.2 Bottom-up categories describe how funding is distributed and why

In Articles A and C, I diverge from the top-down disciplinary categories to assess how funding is distributed across different research topics. The focus in these articles are less on what is funded, but on how funding is distributed and how broad or narrow individual researchers' topic choices are. To do this, I use the Centre for Science and

Technology Studies' micro and meso topic clusters calculated for articles in the Web of Science to represent topics and research specialities. The micro clusters are similar to the algorithmic clusters used in the Leiden Ranking as mentioned above and cover 3,999 topic sets in my data. In addition, I use a set of 865 meso-level topics, which are more aggregate topic groups based on the existing micro topics such that each micro-level topic fits into a meso-level topic. These meso-level topics are used to represent specialities.

The bottom-up topics have the distinct advantage of being much more detailed than top-down disciplinary categories, which is important in e.g. Article C where I investigate researchers' propensity to shift topics. It is likely that most researchers stay within the confines of their discipline throughout their career, while shifting topics according to interests, and both internal and external pressures. Also, the categories are based off researchers' citation behaviour and should be better suited to pick up on behavioural changes in the light of new funding opportunities.

Article D cuts across both the top-down and bottom-up approaches by using both Web of Science subject headings and CWTS' micro- and meso-level topics to discern how funding distributions can be impacted differently across the levels of topical aggregation. However, the use of bottom-up topic categories limits the direct interpretability of each single topic to instead focus on the spread of funding across topics. Figure 4.2 shows a two-dimensional map of the CWTS meso-level clusters. Despite not having direct labels pertaining to their content, we can interpret their meaning through the distinct terms used in the articles within each topic. Three of the most relevant terms are extracted from the titles and abstracts of each topic clusters' publication (Waltman & van Eck, 2012, p. 2383). Each bubble represents a topic and the size reflects the amount of funding for that topic in Denmark between 2005 and 2016.

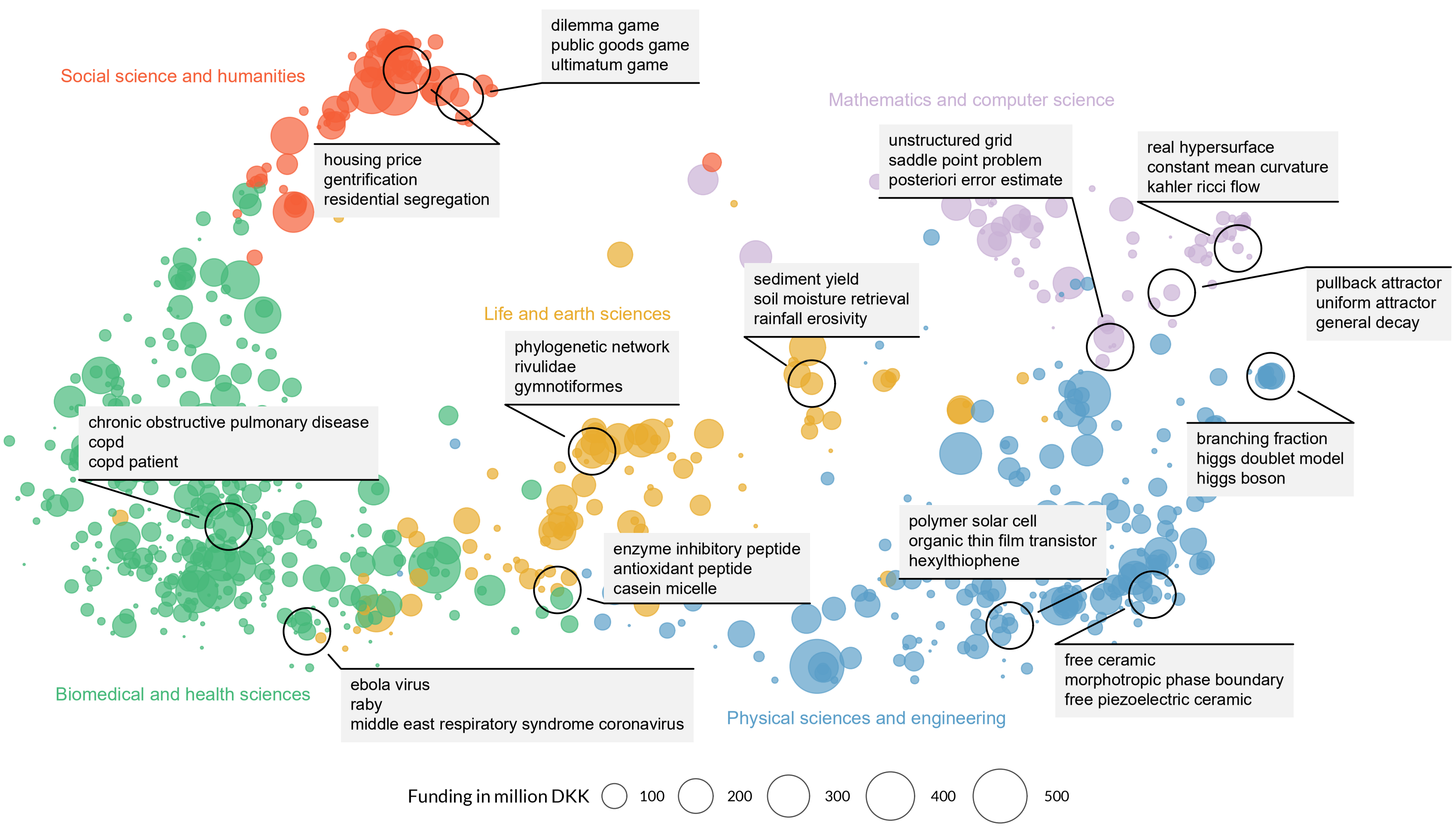


Figure 4.2: Map of funded meso topics in Denmark, 2005-2016.

5. Main findings

In this chapter, I provide an overview of the central findings from each of the four articles within the dissertation. I begin by discussing a central tenet of funding concentration debates: How funding is distributed across individual researchers, research topics, and disciplines. I then turn to the question of how the internal institutional setup of the scientific system shapes these distributions by inducing a static allocation of resources. Finally, I turn to the question of how external priorities by both public and private funders serve to shape funding distributions through an altering of scientific norms and institutions already inbuilt in the scientific system. The aim is here to present a shorter, simplified, and concise presentations of the empirical evidence laid out in the individual research articles, and show how these cut across and contribute to answering the overall research questions.

5.1 Concentration and the distribution of funding: Descriptive evidence

In the following, I discuss how funding allocates across topics, specialities, disciplines, and individual researchers. The focus is placed on the cross-sectional evidence of concentration, accompanied by descriptions of what disciplines and specialities (namely disease areas) are distinctly well-funded. I also provide examples on how funding is concentrated on a select group of individual researchers. The patterns explored here serves as a basis for answering the first research question, and as starting point for the discussions provided in chapter 6. However, it is important to note that almost all of the accompanying papers touch upon this question, and provide some pieces of evidence.

5.1.1 The allocation of funding accommodates a selected group of topics and disciplines

While many papers have shown that funding allocations across individuals follow very skewed distributions (e.g. Katz & Matter, 2020; Larivière et al., 2010; Mongeon et al., 2016), this is has not been established for research content. How are grants and funding amounts allocated across different topics and disciplines? In paper A and B, we try

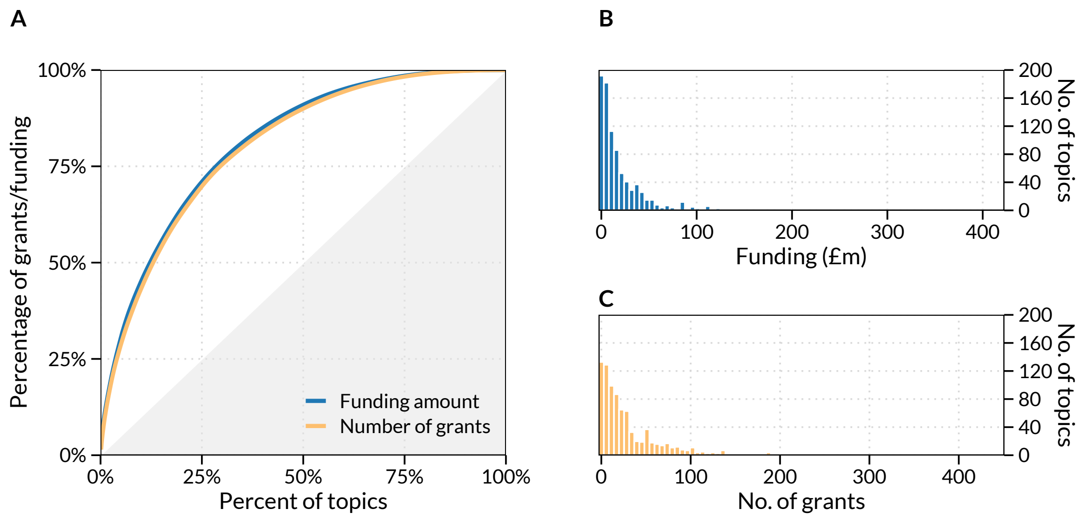


Figure 5.1: Distribution of funding to topics, 2005-2017. Panel (A) shows the cumulative distribution of funding amounts and the number of grants across 858 meso-level topics. Panel (B) shows the distribution of funding amounts across topics. Panel (C) shows the distribution of the number of grants across topics. Based on 30,622 grants awarded by seven research councils in the United Kingdom. The figure is a reprint from paper A.

to answer this question in multiple ways through a wide-gauge analysis of 15 funding organisations in Denmark and 7 research councils in the United Kingdom. Both articles includes a much wider selection of funders and research areas than hitherto, and provides system-level analyses of funding concentration.

Figure 5.1 shows some of the distributional results from paper A, which investigates the spread across 858 meso-level topics described in section 4.3.2.2. Both the histograms and the cumulative distributions show considerable skew in which topics receive the bulk of grants and funding amounts. 25 % of all the topics (around 215 topics) received 70 % of both grants and funding amounts, while 10 % (or 86 topics) were covered by almost 44 %.

The degree of concentration were even more pronounced if funding was analysed on a research council basis. In the Economic and Social Science Research council, 82 % of funding amounts was awarded to 25 % of topics, while this was the case for 77 % in the Engineering and Physical Sciences Research Council. This higher degree of concentration is not surprising, as councils will likely focus on a more narrow set of topics within their designated research area, but also it suggests that councils do tend to cross-subsidise some of the same topics.

Taking both a broader and narrower perspective, evidence from Denmark (in paper B) suggests similar strong concentration of funding amounts. Out of 39 disciplines, only seven amassed 70 % of all funding distributed by public and private funders in 2004-

2016. An overwhelming majority of funds were directed towards biological science and clinical medicine. A more fine-grained analysis of disciplinary concentration in article D (based on 250 Web of Science subject categories) show that it is indeed the biological disciplines with close ties to medical science, biochemistry and molecular biology, which are highly prioritised by all Danish funding organisations. A narrow view within medical science from article B, specifically disease-related medical research, corroborate the overall trend. Out of 134 disease areas, 11 % of funding is directed towards diabetes mellitus and 7 % towards breast cancer. These examples show that competitive research funding is strongly concentrated in specific research content. Using different conceptualisations and operationalisations of content, both topics, specialities, and disciplines, a reoccurring pattern emerges.

Overall, the accumulated evidence from article A, B, and D suggests that funding distributions across different levels of research content (i.e. topics, disciplines, disease areas) follow one of the central expectations for cumulative advantage in scientific recognition. A small segment of content categories accumulates the vast majority of attention from research funders.

This is an important finding for at least two reasons. First, it confirms a first step in the critique of science as a self-organising system. If researchers could freely choose what to study, this would likely result in a skewed distribution of attention. Some topics are just more important than others. However, the degree of concentration found here is much stronger than the moderate degree of funding inequality that might occur naturally (Madsen & Aagaard, 2020, p. 18). Second, the results shows that a strong degree of funding concentration has consequences for the type of research conducted, and a greater reliance on competitive funding in science may bode for a general narrowing of the type of research conducted.

5.1.2 Funding is strongly concentrated in the hands of few researchers

In paper B, we also dig deeper into the question of how funding concentrates by simultaneously analysing the degree of concentration across funded researchers and funded research content. Many studies have found strong concentration of funding without considering both allocations across individuals and the content within funded projects (Katz & Matter, 2020; Mongeon et al., 2016; Wahls, 2018b). We argue that concentration in both domains constitutes a double challenge to the diversity of research systems. Dual concentration of funding limits both the type of successful researcher and research – and by extension – a broad array of scientific perspectives.

We start out by evaluating the degree of funding concentration among researchers who won at least one grant from any of the 15 Danish research funders during the 2004-2016 period. Figure 5.2 shows the cumulative distribution of funding amounts ordered by the size of cumulative amounts per individual researchers. Across 7,539 Danish grantees, the most successful 20 % accumulated 75 % of the total funding sum during the period, while the remaining grantees amassed the left-over 25 %. This is indeed a marked degree of concentration, but does not take into account the many researchers who might have applied for funding during the period, but did not win any grants. We therefore estimated the likely population of eligible researchers to account for the selection bias in our grant data. The most conservative estimate of 15,000 eligible researchers yield a much stronger degree of inequality in funding, where 20 % of potential grantees won 90 % of the total sum of grant money.

Moreover, the strong concentration also reveals some interesting facts about the composition of the select group of grantees with high levels of grant funding. The mean size of grants are also strongly concentrated within the top. Among the top 100 grantees, the mean award size is 90 million kroner and only 30 million for the next 400 grantees. This is not just a case of a few researchers winning one or two very large grants. More often, well-funded researchers will combine one large grant aimed at establishing a research centre, with many smaller grants for individual projects. Among the top-funded, each grantee receive an average of seven grants. Similarly, the gender diversity among top funded grantees is very limited. The top 100 of all grantees also comprise only 16 women, and only 56 female researchers are included in the group of 500 most funded. It indicates a clear case of gender bias in competitive research funding when nearly 40 % of the Danish research population is female.

Just as funding is strongly concentrated within few discipline and topics, the corresponding allocation to individual researchers is highly skewed. A strong implication of this is the lack of diversity in who and what is funded. This is especially important as strong concentration is often argued to have detrimental effects to the furthering of science. A lack of diversity in approaches, methods, and type of question asked may hamper researcher performance, bias against originality and novelty, or limit the relevance of the science conducted to solve societal problems (Aagaard et al., 2020, p. 21)

5.2 The positive feedback of funding allocations

Whereas the results above establish the fundamental basis for studying feedback mechanisms in funding – how funding is distributed – inequality alone does not sufficiently indicate how these distributions emerge and reproduce. A second important part of positive feedback processes in science is the dynamic development of inequalities over time.

Crucially, unhindered positive feedback to initial funding levels should make the well-off increasingly richer, while making the bottom of the distributions relative worse

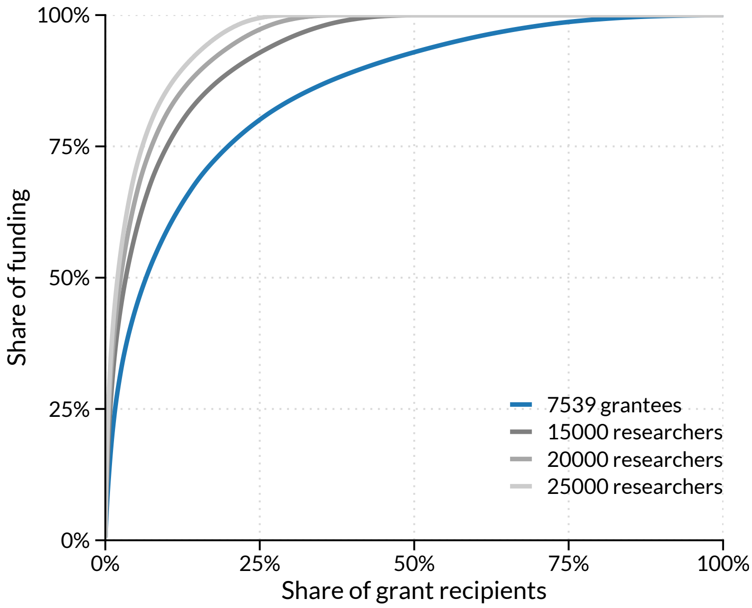


Figure 5.2: Cumulative share of funding for grantees and estimated population of researchers. Gray lines show the cumulative share of funding with an estimated number of non-funded researchers added to the empirical cumulative distribution. The three estimated researcher populations are based on the number of publishing authors from a Danish institution in the period. Based on 19,399 grants awarded by 15 Danish research funding organisations from 2004 to 2016. The figure is a reprint from paper B. Similar figures have also appeared in Aagaard (2019) and Aagaard et al. (2019).

off (Merton, 1968; Price, 1976). In situations where both negative and positive feedback may counteract each other, we should instead expect a static distribution of funding, but not increasing inequality. In essence, initial differences in in the funding of different topics should stabilise or increase as time passes, depending on the mechanisms at work (DiPrete & Eirich, 2006, p. 275). Is this the case? This is an important focus in studying how internal, scientific, institutions of recognition can reproduce strong concentration, and form a large part of the empirical evidence from paper A.

5.2.1 Funding distributions tend to be static and path dependent

In paper A, I investigate how funding distributions evolve over time with a focus on meso-level topical clusters based on citation relationships. For each topic funded through the UK research councils, I calculated the funding percentile of each topic the year it was first funded in the period 2006-2017. I then calculated the funding percentile for individual research topics in the subsequent five years. Topics were grouped into four equal-sized bins (Top 25 %, 50-75 %, 25-50%, and Bottom 25 %) of 214-215 topics each, corresponding to their place in the funding distribution in the initial year. This allows us to track how topics from different parts of the empirical distribution of funding amounts changed or maintained this placement in the following years.

Figure 5.3(a) shows the median and distribution of subsequent funding ranks within each of the four groups. Within the top 25 %, topics have a tendency to stay within the top of the funding distribution. The median funding rank is the 71st percentile one year after, which increases slightly to the 73rd percentile in the four years after. The topics starting out in the bottom 25 % tend instead to remain around the 15th percentile in the following five years. These developments indicate a great deal of stability in initial funding allocations. All topics tend to decrease a small amount in funding rank in the first year, but then stabilise. Looking at the actual monetary value allocated towards each topic, the path dependency of funding distributions seem to replicate somewhat.

In figure 5.3(b), I show how the predicted levels of funding allocated to a topic varies with prior amounts of funding in one to five years before, based on a hierarchical hurdle-gamma model. The model indicates some positive feedback from previous funding. Previous funding of about five million pounds in one to four years prior approximately translates into an average of around 1 million pounds of funding in a given year for a topic. Importantly, larger returns to previous funding are more pronounced for higher levels, indicating increasing returns to prior funding as the prior amount grows.

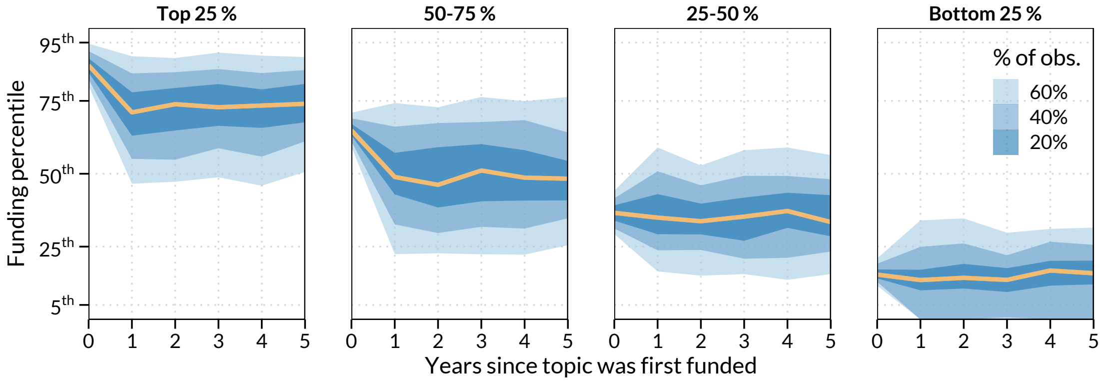
Conversely, the amount of funding for at topic five years before does not seem to matter much, and show no increasing returns. Instead, higher levels of prior funding tends to correspond with slightly lower levels of current funding. One interpretation of this could be that path dependency for research topics does not continue indefinitely, but instead manifests as shorter cycles of popularity. This is congruent with a dynamic research system where some topics are heavily prioritised for a period, and working within these entails a greater probability of funding, while then levelling off as new topics take over.

5.2.2 Concentration for individual researchers and topics are interdependent

In chapter 3 and section 5.1.2 above, I claimed that positive feedback for some topics is closely connected to the high degree of concentration in who receives funding. In paper D, I argue that the distributions of funding to topics and individual researchers are heavily interdependent because researchers within certain areas are much more likely to receive funding. Consequently, some topics receive more attention because a small group of researchers acquire the bulk of competitive funding. In essence, this is a self-reinforcing relationship, with no obvious beginning or ending. The question is if individual grantees in the top of the funding distribution are more likely to focus on well-funded topics?

At first glance, the answer to this question may seem slightly tautological. Admittedly, topics are not funded per se. Only people receive funding for specific projects. Stratification in which topics and disciplines are funded therefore crucially depend

(a) Subsequent funding percentiles for topics.



(b) Average funding amounts for topics conditional on previous funding amounts.

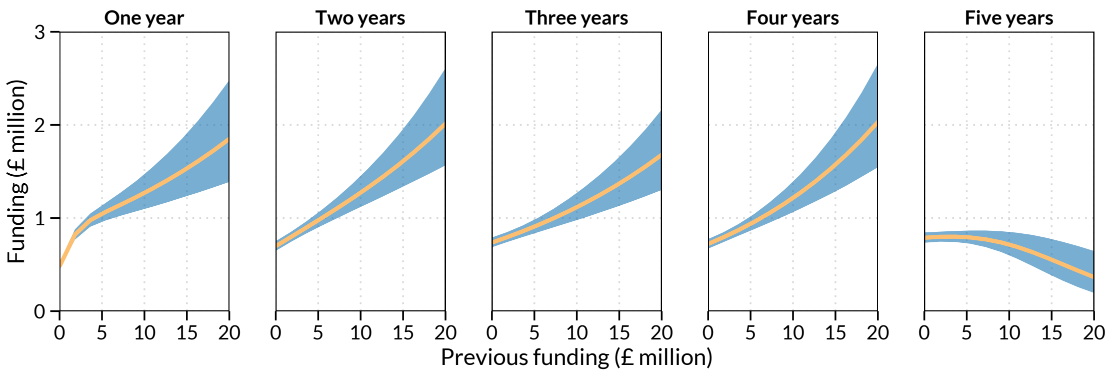


Figure 5.3: Stasis in funding distributions. Panel (a) shows the median funding percentile for meso-level topics depending on their funding percentile when first funded. Each group correspond to 25 % increments of the funding distribution when topics were first funded. Panel (b) shows the posterior predicted level of funding for a topic dependent on previous yearly funding levels in one to five years before. Intervals are 95 % credible intervals from the posterior distribution of a hierarchical hurdle-gamma model. Both figures are reprinted from paper A.

on stratification in who receives funding, and that those well-off individuals tend to research a narrow selection of topics or are active within only a handful of disciplines. However, we could imagine that topics receive more attention from funders from another perspective. It may instead be that most researchers – well-funded or not – simply focus on the same type of research content. This would be congruent with a self-organising system of research, where some topics appear more important to all who partake.

In paper D, I make a first attempt at untangling this possible interdependence. In the article, I investigate who are more likely to receive funding for the most popular topics and disciplines. In figure 5.4, I place each grantee within their funding percentile and aggregate these in 20 groups spanning five percentiles (i.e., the bottom 5 %, 5-10 %, etc.).

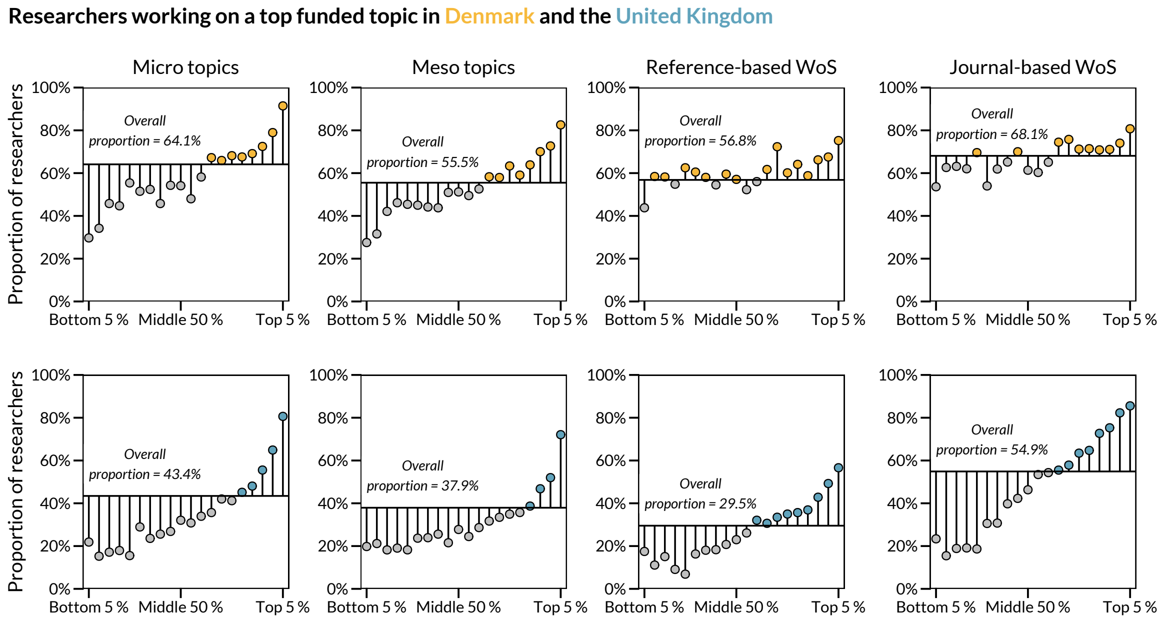


Figure 5.4: Individual funding concentration and propensity to work in top 5 % funded research fields. Proportion of grantees within each group whose funded project(s) addressed a top 5 % topic or discipline. The figure is reprinted from paper D.

I proceed to calculate the proportion of grantees in each group who received funding for research in the top 5 % of funded topics and disciplines for both Denmark and the United Kingdom.

In Denmark, there is an overall tendency to work within disciplines and on topics that are within the top of the funding distribution. For micro- and meso-level topics, over 50 % of researchers work within the top funded topics. The pattern is similar for the disciplinary level (250 Web of Science categories). In the British data, grantees are less inclined to work on similar topics and in the same disciplines.

For both funding systems, one pattern is fairly consistent. At the micro- and meso-topical level, grantees in opposite parts of the funding distribution differ in their propensity to work on well-funded topics. Looking at e.g. the distribution of attention to different meso-topics, 80 % of top funded researchers work on the most funded topics. Similarly, only 20-30 % of the grantees in the bottom of the distribution work on the most popular topics. This stratification in who works on the most funded topics then seems to be an important mechanism linking individual and topical/disciplinary concentration of funding.

5.3 How external priorities shape funding allocations

So far, this chapter has been preoccupied with questions of how funding is distributed and how competition within the scientific system affects the way funding allocations evolve or remain static. Now, I turn the lens towards possible sources of outside influence over what is funded, and how the priorities of government research councils and private research foundations shape both the system-wide distribution of research grants and the individual behaviour of the grantees.

5.3.1 Thematic priorities in public funding competitions

I have argued that the increasing use of thematic priorities, or targeted funding, could have the potential to shift both the internal competition and norms of the science system in general, but also shift the individual incentives faced by each researcher when choosing topics to study.

First, I specifically hone in on the individual-level effects of targeted funding towards particular research content at the topical level. A prevalent view in the science policy debate is that targeted funding either increase the incentives for researchers to engage in prioritised topics, or lead them to recast their existing research lines as relevant to these topics. In paper C, we use a matching design to compare researchers in two British research councils. One group received regular project funding from competitions without any topics prioritised, and another group addressed one or more priorities in their project proposals. Specifically, we ask how different their topical focus were between research conducted before they received a grant, research they conducted as part of a funded project, and research conducted after receiving a grant but not connected to this grant. By comparing recipients of different types of grants, we can investigate if prioritised funding is more likely to shift a researcher's topical focus than regular, responsive mode, funding.

A purely descriptive comparison shows that both the mean and median similarity between previous research and funded research is lower among those who receive prioritised grants. These researchers tend to shift their topical focus more in response to a grant. However, the differences between the groups disappear when comparing previous research to unrelated post-grant research. In these instances, both groups engage in topics that are quite similar to their past work. Nevertheless, we may be a bit sceptical about this simple comparison. Are the two groups even comparable? Researchers working within certain topic areas may never be able to win prioritised funds because their preferred topics are never prioritised. Or perhaps researchers who address topical priorities are more experienced, and have produced a lot more papers in order to risk such an application? This could perhaps also structure how similar their research are, as more productive and experienced researchers likely cover more research lines (see e.g. Horlings & Gurney, 2013).

To ensure greater comparability between the two groups, we matched the individual researchers on a selection of pre-grant factors, including their primary discipline, num-

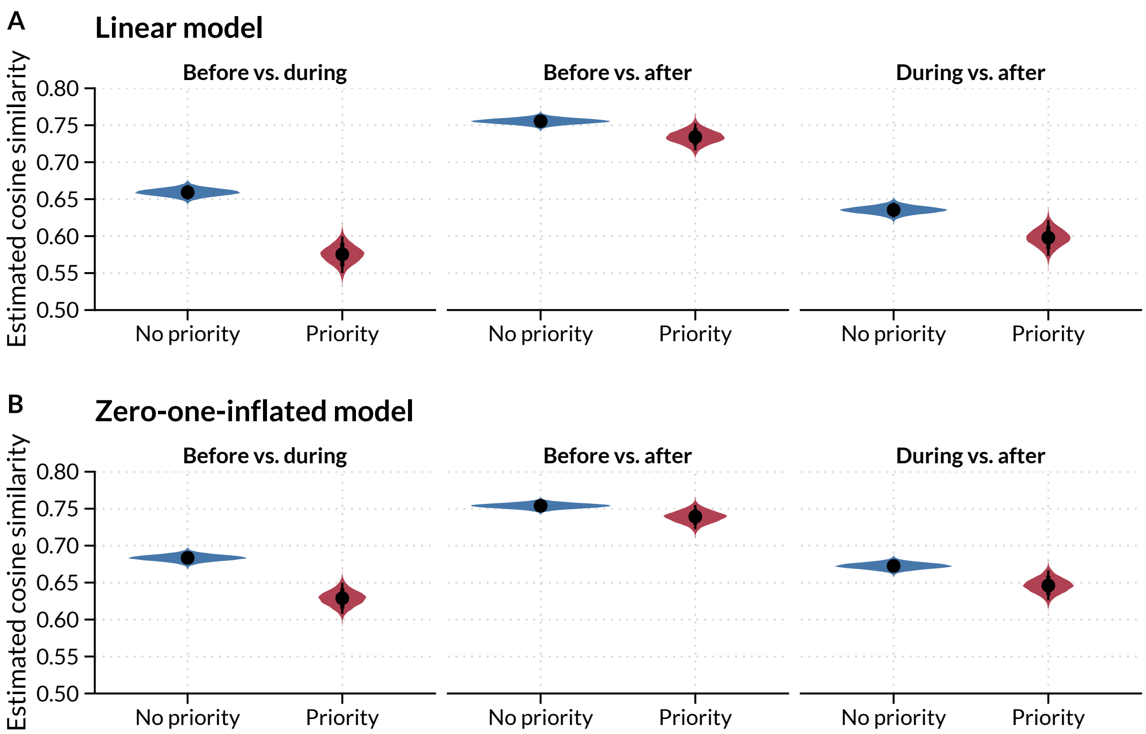


Figure 5.5: Estimated topic similarities for researchers receiving prioritised or non-prioritised funding. (A) Estimated mean similarities derived from the coarsened exact matched sample using a Bayesian linear model. (B) Estimated mean similarities derived from the coarsened exact matched sample using a Bayesian zero-one-inflated beta model. Point estimates are medians of the posterior distributions (coloured violin plots), with 80 % (thick lines) and 95 % (thin lines) credible intervals. N = 5,184. Reprinted from paper C.

ber of published papers, citations, and more. We then used are series of linear and zero-one-inflation regressions to model the difference in similarities between groups. Figure 5.5 presents the posterior distributions and estimates from these models. We estimated separate models for three comparisons: topics investigated before receiving a grant vs. topics investigated in a grant, topics investigated before vs. after a grant, and topics investigated in a grant vs. after a grant.

Two core findings stand out. First, the patterns of difference are very similar to the descriptive comparisons. Recipients of prioritised funding are more likely to engage in new topics within the confines of their grant. However, they subsequently return to research similar to their pre-grant work at the same rate as recipients of non-prioritised funding. Second, we also checked that the grantees in the group of did not engage in more diverse topic portfolios by adding new topics to their existing research lines. This is not the case. Both groups seem to work on a core set of few topics that are already very similar to begin with. In summary, it does not seem like prioritised funding delivers a long-term shift in research focus for individual researchers. Instead, it provides a short-term shock to their research portfolio equilibrium.

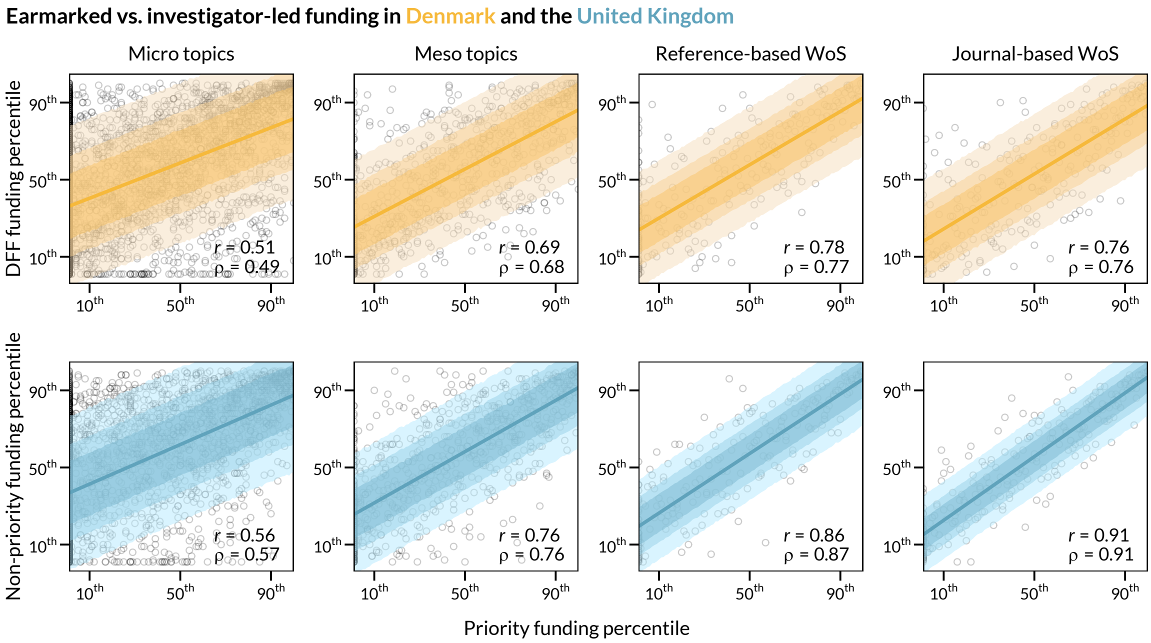


Figure 5.6: Topical overlap of funded priority and non-priority research in Denmark and the United Kingdom. Solid lines are regression fits from a Bayesian linear model, with 50 %, 70 % and 90 % credible intervals calculated from the posterior predictive distribution. r = Pearson's correlation and ρ = Spearman's rank correlation. Reprinted from paper D.

What are the broader effects of prioritised funding across the funding systems? In paper D, I try to assess how targeted/prioritised funding lines up with the type of research conducted in non-targeted grants in general. Here – it seems – targeted funding in both Danish and British research councils show a great overlap at the disciplinary level, while having less of an overlap at the more fine-grained topical level. In Figure 5.6, I have calculated the funding percentile of each topic and discipline within targeted and non-targeted groups of funding. The figure also shows the predicted bivariate relationship between prioritised and non-prioritised funding percentiles. Depending on the exact conceptualisation of disciplines, the funding percentiles for each discipline show a moderate to high correlation (r = [0.76 - 0.91]) in both Denmark and the United Kingdom, indicating that a few disciplinary areas are far more successful in funding competitions of both the targeted and non-targeted variety.

At the topical level – both fine-grained micro-topics and more aggregate meso-topics – the overlap is less extensive. Even if funding percentiles for individual topics are positively correlated, there is much greater variability in both a Danish and British context. So even if targeted and non-targeted funding tends to concentrate within the same overall areas of science, it seems as if there is greater room for specific topic choice within these disciplines.

5.3.2 The spill-over of private funders' priorities

The increased reliance on grants and funding from private research foundations have spurred questions on how the priorities of these foundations shape the incentives to study particular topics. While existing work on how private funding affects the pace and direction of research have used publication data to show a spill-over of privately funded into publicly funded research (Evans et al., 2014; Yegros-Yegros et al., 2020), they have been unable to pin-point how much funding is invested in similar research areas. Rather, scholars have used publication volumes as proxies for investment levels from different sources (Ciarli & Ràfols, 2019, p. 950), but without being able to more exactly gauge how many resources are invested.

In paper B and D, we extend studies conducted with publication data (Evans et al., 2014; Ràfols & Yegros-Yegros, 2018; Waltman et al., 2019; Yegros-Yegros et al., 2020; Yegros-Yegros et al., 2018), and funding-based studies within single disciplines (Head et al., 2013; Head et al., 2016). We examine how public and private funding organisations invest across topics and within disease-related medical research, providing a view across all disciplines and a view from within a discipline of great relevance to private funders. In both paper B and D, we show that funders across the board tend to focus on similar disciplines.

In Denmark, 25 % of the OECD-defined disciplines get 73 % of all funding from public funders, and 86 % from private funders. However, while both types of funders prioritise a small set of areas, some differences in funding portfolios exists. Public funders direct the majority of their grants towards biology, physics, chemistry, and clinical and basic medicine. Private funders instead focus primarily on biology and clinical medicine.

These tendencies become clear when moving to the topical level. Figure 5.7 shows the correlations between public and private funding percentiles across different conceptualisation of topics and disciplines. Again, the overlap is considerably stronger at the disciplinary level, and weaker at the more granular topical level. At these aggregate levels, it is not clear whether private priorities spill over into public funding, or simply that funders from each domain focus on the most important and promising areas. To examine this further, paper B and D delve deeper into the distribution of funding within the biomedical area.

In paper B, we change focus from public and private research investments across all disciplines, and focus on disease-related research. This is an area with great relevance for Denmark's many private research foundations as they base their endowments on ownership of large-scale pharmaceutical and chemical companies such as Novo Nordisk, Lundbeck, Novozymes, and ALK. We compared the funding of 134 disease categories

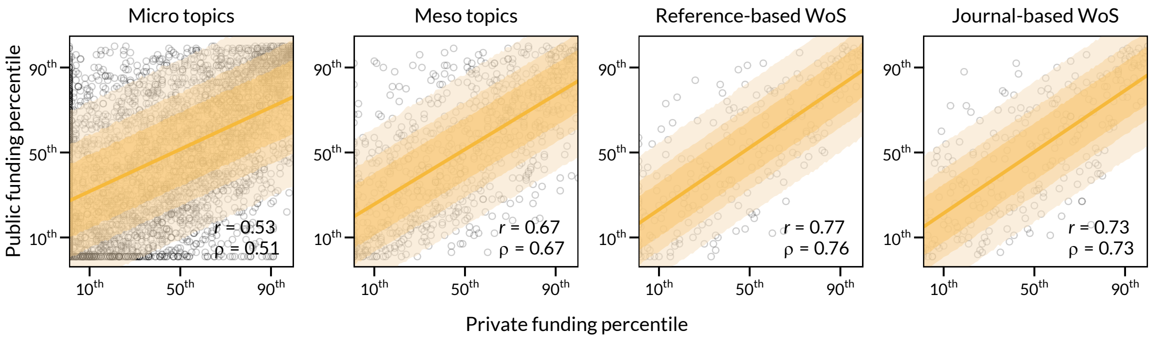


Figure 5.7: Topical overlap of funded private vs. public research in Denmark. Solid lines are regression fits from a Bayesian linear model, with 50 %, 70 % and 90 % credible intervals calculated from the posterior predictive distribution. r = Pearson's correlation and ρ = Spearman's rank correlation. Reprinted from paper D.

with measures of their societal health burden, and how each disease was prioritised in public research councils and private research funding foundations.

First, 4,108 grants had some relevance to at least one disease area. Aggregating the value of these grants across diseases shows a large skew in what is funded by both public and private funders. Some disease areas receive a much larger portion of total competitive funding than their portion of health burden would predict. This is not to say that funding levels should always follow the health burden of a disease, but it seems that some misalignment exists. Diabetes, breast cancer, and skin diseases receive 11, 7, and almost 4 % of total funding, but their relative share of the total health burden is somewhere from 2 to 4 %. Of course, this is the case because Danish funders invest heavily into these areas, but a lot of disease areas with a much heavier burden appears underprioritised. Ischemic heart disease constitutes over 6 % of total disability-adjusted life years lost to disease in the period 2004-2016, but receives under 4 % of funding. Another area – back and neck pain – is the most burdensome condition with around 9 % of total disability adjusted life-years, but receives just short of 1 % of disease-related funding.

The apparent reason for these misalignments is not that public and private funders do not invest in burdensome diseases. Instead, private and public funders invest in the same disease areas amplifying this misalignment to a much larger extent. Figure 5.8 shows the overlap in funding percentiles for each disease, sized by its burden. Following the same logic from figure 5.7 above, the correlation between funding percentiles is substantial. It is not that private and public funders invest the same amounts in different diseases, but that similar diseases are located similarly in public and private funding distributions. As an example, breast cancer was awarded 331.6 million kroner from public funders and 257.7 million kroner from private funders in the period, but because public funders invested more heavily in disease research, this places it within the 98th percentile in public research councils and 99th percentile in private foundations.

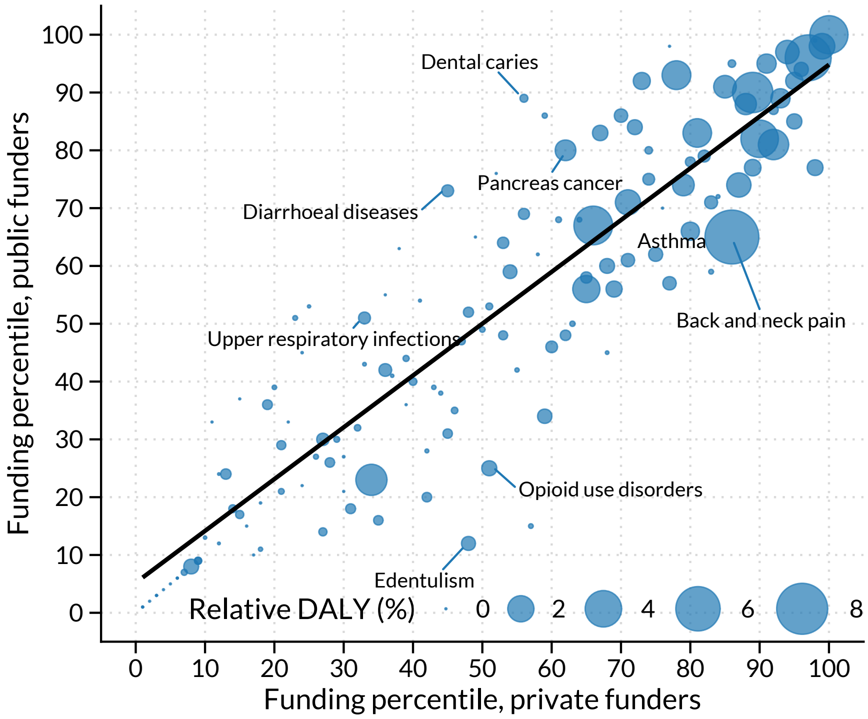


Figure 5.8: Funding percentiles for 134 disease areas in private and public funding bodies in Denmark. Axes show funding percentiles within either all public or private funding bodies, and is based on 4,108 grants in total. 0 denotes the lowest percentile, and 100 denotes the highest. Size of circles is proportional to the relative average disease burden in Denmark, 2004-2016. Reprinted from paper B.

Paper D also estimated the overlap between private and public priorities through a series of hierarchical models taking into account varying overlap between five different research areas: Biomedical and Health Sciences, Life and Earth Sciences, Mathematics and Computer Science, Physical Sciences and Engineering, and the Social Sciences and Humanities. Looking at almost 4,000 micro topics, a one percentile difference in private funding is associated with an average 0.75 public percentile difference in public funding for the biomedical field. Comparing this to the social sciences and humanities, the difference in public funding is only 0.31 percentiles.

In sum, there tends to be substantial spill-over of private funding priorities, but these are localised within disciplines or research areas with the greatest relevance to particular private funders.

6. Concluding discussion

By addressing funding concentration from the perspective of what type of research content is sponsored, the four articles in the dissertation have contributed with a new empirical basis for discussing the consequences of increased reliance on competitive research funding. Contrary to many former studies, the empirical studies presented here focus on how research funding is allocated to different research topics, specialities, and disciplines, rather than on the distribution of funding across individual researchers. This entails a shift in focus toward looking at the consequences of competitive funding for the direction of research. In this dissertation, I have illuminated the degree of funding concentration of topics both at the system level and at the individual level to give a more comprehensive view of the issue. Throughout the different studies, I have also applied different levels of analysis to examine the patterns with different levels of granularity.

Firstly, I have contrasted systemic developments in what is funded across multiple funders with a focus on what individual researchers engage in. A broad focus across the systems illuminate the overall degree of concentration and the role of the funding system in shaping patterns of concentration. Further investigating what individual grantees choose to study instead serves to explore how researchers potentially react to incentives inbuilt in the system, and the external priorities shifting these incentives.

Secondly, I have analysed the distribution of funding across research content at different levels of detail. The disciplinary level highlights broad tendencies and points towards how funding is prioritised by both political and private actors in the research funding system. The topic and speciality level, on the other hand, show more fine-grained details. Such a focus on more restricted topics helps indicate how internal dynamics may shift or maintain the focus of the individual researcher, and how funding concentration can play out within disciplinary boundaries. Throughout the different articles, I have also attempted to empirically explore potential drivers of concentration. Where former studies have mostly been occupied with establishing some of the deficiencies of the funding system, my contribution also seeks to explore and discuss the drivers of this development. This is necessary to identify potential causes and to enable us to remedy and address issues that many researchers perceive to be problematic for knowledge generation. In sum, these different considerations contribute to an overall discussion of how internal competition and external priorities can shape the distribution of research funding across research topics.

In this final chapter, I take stock of the different empirical results to discuss how internal and external drivers of concentration combine to shape funding distributions. I then return to the implications of these findings by asking whether research concentration entails positive or negative outcomes for individual researchers and the scientific system more broadly.

In the end, the work presented here is still a first, imperfect, step in empirically investigating how funding shapes research content. The third section of this chapter therefore highlights some strong limitations of my approaches and presents some directions for future research. I conclude by discussing what the results implies for the future of the competitive research funding system, and what lessons emerge with direct relevance for funding policies.

6.1 A complicated tale of multiple, interrelated drivers

I initiated chapter 1 by highlighting how studies of research funding have been preoccupied with studying its allocation across researchers and how funding concentration entail both positive and negative career outcomes. I also argued that discussing how funding concentration leads to more or less publications and citations could be the wrong yardstick for judging the impact of increased reliance on competitive funding. Instead, we should turn to questions of *what* type of research is funded and *how competitive research funding is distributed across research content*.

Studying funding distributions across the Danish funding landscape and in the British research councils confirms a rather consistent pattern. Competitive research funding is concentrated in the hands of a small group of grantees, and centred on a relatively narrow selection of research topics or wider specialities and disciplines. Across both countries, around 20 % of researchers accumulate 70-80 % of the total funding amounts, the most well-funded researchers amass very large pools of resources, and are generally awarded grants with higher monetary value that other grantees. When analysing the spread of funding across categories of research content, similar patterns of concentration arise. Across the Danish funding system, few, very similar biological and biomedical disciplines attract around 70 % of funding amounts. Similar patterns of funding concentration is evident within disciplines. In paper A, I showed that the 20/70 distribution is prevalent within individual research councils, and paper B highlighted how a few disease specialities dominate funding competitions in biomedical research. The latter also showed that these distributions do not necessarily correlate well with measures of societal needs.

The consistent patterns of strong funding concentration also led me ask: *What are the potential drivers of research funding concentration or dispersal across research content at different levels of granularity?*

In chapter 3, I posited that the personal incentives of each individual researcher is the driving force for choosing different topics or disciplines to study, and ultimately apply for funding. These incentives are complex and changing, but are firmly connected to how we as researchers perceive future opportunities, prestige, intrinsic value, and perhaps societal needs. It is the individual weighing of the different incentives that underlies topical and disciplinary choices, and these choices aggregate to form a distribution of funding across topics. How funding is distributed thus depends on what individual researchers choose to study, what type of project they seek funding for, and what ideas the community (especially those reviewing applications) find worthy of funding.

However, such a simple science-internal model for understanding both individual behaviour and systemic structures has never been an accurate description of the scientific process. Science is not conducted in a vacuum, where only the opinion and work of other scientists matter. How incentives for studying certain topics emerge are a complex process of many different internal and external factors. I therefore proposed to focus on at least two broad factors with the potential for changing incentives for studying certain research content, and for how it is evaluated by the scientific community.

One argument has been that the structures for assigning and allocating rewards and recognition *internally* in the scientific system may perpetuate a skewed focus on certain researchers, research topics, specialities, and disciplines. The direct effects of a differential distribution of recognition on what and who are funded are however hard to entangle based on grant data. Merton and Zuckerman (Merton, 1957; Merton, 1988; Zuckerman, 1967; Zuckerman, 1970) originally build the case for the Matthew Effect using detailed interviews with American Nobel Laureates, who themselves admitted to experiencing the positive reinforcement and feedback that their status provided them with compared to their peers. Instead, I have resorted to investigate some of the more indirect indications of how institutions of peer-review may influence the concentration and dispersal of funding. A consistent finding in both the Danish and British context, is the tendency for funding to be concentrated on a small elite of researchers and few topics, specialities, and disciplines. It is a pattern, which repeats itself across and within private and public funding domains, but also within individual funding institutions such as the UK Research Councils. These, more or less, cross-sectional patterns confirm that funding is concentrated, but do not necessarily imply that mechanisms of positive feedback are at work. However, at least two of the results stemming from this dissertation appear to indicate that some level of positive feedback could be present.

Firstly, Paper A shows that the distribution of funding in UK Research Councils tends to be very static over time. In the short run, the same topics and specialities figures in both the top and bottom of the distribution in a given year. This is especially true if we compare the very top of the distributions (e.g. the 90th percentile and above) with the bottom 10 %. In both instances, topics overwhelmingly stay within their initial part of the distribution. This pattern is strongest for biomedical, physical, and biological sciences (i.e. in the MRC, BBSRC, and the EPSRC), while funding distributions seem more fluctuant in the humanities and social sciences (cf. AHRC and ESRC). Such differences could likely be due to the stronger dependence on external funding in the "harder" sciences, where competition for funds is more intense, and large scale funding is often necessary to carry out a project. Conversely, the paper also shows a clear limit to the positive funding feedback as this effect seems to wane off with time, such that funding levels five years past seem to have little impact on current levels. Hence, positive feedback seems to reinforce a skewed distributions of funding in the shorter term, but some negative feedback mechanisms appear to set in in the long run. This is congruent with the expectation that both feedback effects often will work in unison to create some level of path dependency[1] in what is funded. These empirical patterns are only shown for the British data, focused on responsive mode grants where we would expect internal norms to have the highest impact. While the broader generalisability of these finding may appear somewhat lower, it shows that shorter-term path dependency is present in funding competitions where pure internal norms of recognition and rewards are the primary levers for evaluation in grant reviews.

Secondly, Paper D provides some descriptive evidence for the interdependence of individual level concentration and concentration at the topic, speciality and disciplinary level. A strong concentration within certain areas of science, be it topics or disciplines, could equally indicate that the funding system is self-organising. Skewed patterns of funding would then just reflect differences in internal demand for different topics or disciplines as these differ in opportunity for discoveries. This would likely lead a large share of grant applicants to address these topics. However, this is not the actual empirical pattern. The most well-funded topics are predominantly addressed by researchers in the top of the funding distribution (and vice versa) and suggests that individual level concentration can have adverse effects on the diversity of research content funded. In both Denmark and the UK, this differentiation is prevalent at the micro- and meso topic

[1]The reader may disagree that such short-term "stickiness" of funding distributions can credibly be called path dependency. While I agree that this constitutes conceptual stretching of the term, I think it neatly describes the pattern that funding levels Y_{it} for a "topic" i in year t depends on funding levels for Y_{it-k}, where k is some time lag of e.g. a year, two years, etc.

level. At the disciplinary level, there is a much more equal spread of disciplinary choice across researchers in Denmark, but not in the UK. This could be due to the differences in the sample of funders analysed, where the British data does not include private foundations or charities, which may even out the differences.

Admittedly, some level of self-organisation could also affect the distribution of funding across research content, but I have not conducted any explicit analysis of the internal demand in the scientific community. How funding distributions align with what is published could provide a way to investigate this. This has explicitly been addressed by Klavans and Boyack (2017a) as a measure of the internal demand for specific topics, showing that a third of funding variation can probably be attributed to variation in demand or topic prominence in the literature. They find that highly prominent topics receive far more funding per researcher than topics that are not prominent, and note that this does not necessarily represent the breadth of demand that is associated with societal goals (Klavans & Boyack, 2017a, p. 1171). On the other hand, both past prominence of topics in the scientific literature and subsequent funding levels likely reflect funding distributions in the past, as publication and funding patterns often correlate (Ciarli & Ràfols, 2019, p. 950).

While the dissertation's analyses of the internal drivers of funding concentration have been rather indirect, a second focus on external drivers have permitted some more direct explorations of how funder priorities may structure what type of research is funded and studied. With respect to government funding priorities, these largely target already well-funded disciplines within the medical and natural sciences in the two cases I have analysed. This may reinforce concentration at the topical level. The results from Article D shows that the overlap between responsive mode funding portfolios and funding portfolios in strategic research councils or targeted grants are greatest at the disciplinary level. Some degree of overlap exists at the speciality (meso-topic) level, while portfolios are much more diverse at the topic level. Comparing the Danish and British case shows that overlaps are greatest in the UK, where overlap at topic and speciality levels is also prominent. It seems likely that this difference pertains to how specific a priority is formulated. Priority grants from the UK Research Councils[2] were often formulated within a council and on more specific topics (e.g. antimicrobial resistance), while Danish government priorities were enacted through broader strategic research councils. The large degree of overlap seems to confirm some interaction between internal and external drivers of funding concentration. However, the nature of this interaction remain obscured. Concentration patterns could be due to researcher involvement in setting priority topics following trends within the scientific community, or due to priority topics becoming prevalent in responsive mode funding as well.

[2] Although, cross-council priority areas are also prevalent.

Article C and D is an attempt to discern how this latter point may play out at the researcher level. A comparison of researchers receiving responsive or targeted grants show a limited sustained influence or priority funding in the British context. The results show that:

- Public funder priorities affect the topic choices of individual researchers in the short run. Recipients of thematically targeted grants publish research topically more different from their earlier work compared to other recipients.
- In the longer run, this difference disappears when comparing publications patterns before and after their grants.
- Topical similarity is lowest in the years prior to receiving a targeted grant, while their work 1-4 years after is more similar.

This is more consistent with researchers applying for funding within their broader field of work, and targeted funds perhaps temporarily shifting what they work on is largely in line with research on targeted grants in the U.S. National Institutes of Health (Myers, 2020). It lends support to the notion that targeted funds can primarily function as a temporary way of filling a knowledge gap, but probably do not induce major shifts in what researchers are studying.

Focusing on private research foundations in Denmark has also served to illuminate how the priorities of private funders interact with internally formulated priorities in responsive mode funding. Examining the type of research funded shows that certain areas within the biological and biomedical sciences in Denmark are extremely well funded from both private and public sources:

- There is a large overlap in what type of disease-related research is funded and in the disciplines funded through public and private funding (e.g. biochemistry and molecular biology, endocrinology and metabolism, and oncology).
- This is likely not due to certain diseases exerting a higher health burden as funding and burden levels are not highly correlated.
- Private funders are more focused on disciplines such as psychiatry and genetics, whereas public funders have high pay-outs for plant biology and immunology.
- At the individual level, researchers receiving the most grants often receive funding from both public and private means.

Taken together, the overlap in what and who is funded suggests a strong interaction of internal scientific norms and private funder priorities. The latter point exemplifies that prestigious privately funded grants often lead to subsequent grants from public funders (Hallonsten & Hugander, 2014, p. 255). Such dynamics may create an overlap in what is funded because well-funded researcher rely on both public and private funding, and likely experience a lock in of problem choices. This overlap could stem from the very similar criteria on which most research funders evaluate grant proposals. The majority of funders aim to support the best researchers by using quality criteria in the form of publication histories, prior grants, and citation performance. This practice may be sound at the individual grant level, but together create an unintended degree of concentration as funders lack oversight of the broader funding portfolio (Aagaard, 2019). The result presented here thus suggests that fragmented funding systems with many competing private and public funders are prone to some unintended, systemic, concentration effects.

In sum, I have illuminated that a strong degree of funding concentration for research content can be rooted in shorter fortuitous cycles in funding, coupled with strong interdependence of individual and content level concentration, and some effect of priorities formulated by funders. Overall, both the funding systems in Denmark and the UK show signs of path dependency in what is funded, with indications of internal norms and external priorities sometimes perpetuating, sometimes counter-working, each other.

6.2 Are high degrees of funding concentration and path dependency problematic?

Policy debates and empirical studies often portray strong concentration of funding as detrimental or damaging to the conduct of research and the functioning of the science system as a whole. I have also myself implied this throughout the dissertation. However, this is obviously a matter of degree and we should ask ourselves whether the current levels of funding concentration are in fact counterproductive or not, and if greater concentration or more funding dispersal would change our conclusions.

To have a nuanced discussion of these issues it is important to recall that some perspectives also emphasize a set of positive effects of bundling resources as a counter thesis to the call for greater dispersion of funds. Concentration of funding may not be inherently bad.

The question is often portrayed as two opposing views: Are scientific advancements driven by a small, high performing, elite, or is it the accumulation of a mass of mediocre and narrowly specialised work? Cole and Cole (1972) note that most citations in physics are produced by a small proportion of physicists, that most highly cited work mostly refer to other highly cited work, and that most "eminent" scholars are trained in few select departments. Bornmann et al. (2010) corroborate this across both the life, health, physical, and social sciences. Most high and medium cited papers rely on the most cited research to an equal degree (Bornmann et al., 2010). On this background, arguments in favour of strong funding concentration often invoke the observation that the actual distribution of talent is very skewed, and current funding allocations are more equal than talent differences would entail (Hicks & Katz, 2011). Concentration is then a consequence of self-selective bias in allocation of grants, where those who are talented and motivated also apply to a much higher degree (Bloch et al., 2014, p. 81). Selectivity and concentration are therefore necessary to retain or motivate researchers and provide clear incentives to excel (Vaesen & Katzav, 2017, pp. 8–9).

Counterarguments highlight that talent may be genuinely skewed, but do not manifest itself in observable differences in performance. We might not even have reliable proxies for researcher skill or talent and empirical investigations of this may rest on faulty assumptions. Arguments for a strong concentration of talent often rely on measures of scholarly impact, such as citations. But citations are likely a flawed representation of actual skill or talent of a researcher and her research. Whereas citations may convey the impact of a piece of research, and by extension the researcher, references do not linearly correlate with quality (Nicolaisen, 2007, p. 632). Instead, large numbers of citations may be given to work of either high or low quality for different reasons. Low quality science may be highly cited because authors seek to criticise it, indicate it as a vital background reference, note a controversy, or agree with it (Garfield, 1978). The relationship between quality and citation accumulation is therefore likely a J-shaped correlation, where both high and low quality research exhibit higher impact that work of medium quality (Nicolaisen, 2002). Citation counts are also imperfect measures because a lot of influential work remains uncited. MacRoberts and MacRoberts (1986; 1987; 2010) argue that evidence of strongly skewed distributions of talent and influence in scholarly work is largely unfounded because researchers tend to cite work in order to persuade rather than give credit. Consequently, a lot of influential work is seldom referenced, while subsequent work building on this may receive the bulk of citations (McMahan & McFarland, 2021). Finally, citation accumulation in the short and long term could signify different things. Short term citations indicate engagement with current literature and discourse, while long term citations could indicate real impact or sustained visibility (Leydesdorff & Bornmann, 2016). In conclusion, it may be that talent, skill, or importance of individual researchers and research is distributed in an unequal fashion, and should merit concentration of funding resources, but a solid empirical foundation of such claims seems to be lacking.

Concentration of funding as a natural extension of differences in talent or effort furthermore rests on the assumption that the distribution of funding is decided upon in an unbiased and straightforward, meritocratic, way. Given that most competitive research funding is allocated on the basis of peer review, this could boil down to a question of how well reviewers of grant applications can identify projects with the most potential. Given some of the arguments I presented in chapter 3, we may initially be sceptical of this.

To my knowledge, no studies have made an unequivocal comparison of the scientific potential or importance stemming from funded and unfunded grant applications. Plenty of analyses have tried to identify whether higher peer-review scores in grant competitions lead to more citations. The evidence seems to be mixed and not entirely conclusive. It has been suggested that a one standard deviation worse peer review score is associated with an average 7.3 % fewer future publications and 14,8 % fewer citations when comparing grant applicants within the same year, with similar academic background, experience, and publication record (D. Li & Agha, 2015). Similar findings have been reported for peer review processes across different research funders in both Europe (Bornmann & Daniel, 2006; Bornmann et al., 2008) and the U.S. (Gallo et al., 2014), albeit in much smaller samples of grant applications. Applicants selected by peer review panels, or receiving higher review scores, tend to be more productive and have higher citation impact in their subsequent career (Gallo & Glisson, 2018).

However, these results also leads us to question if citation measures reflect the importance of a funded project. The relationship between peer review scores and project outcomes may well be confounded by unobservable characteristics affecting both the quality of a proposal and the applicant's later productivity and cumulative citations (for discussions see Danthi et al., 2015, p. 786; Park et al., 2015, p. 1153). The effects of higher review scores on project outcomes tend to be variable. Better scores may lead to higher average productivity and impact, but many studies find that bibliometric indicators vary immensely between projects with similar scores (Fang et al., 2016; Gallo et al., 2014). Further, higher citation impact may not be a desirable outcome of a grant. The publication process is marred by random chance, unpredictability, and extreme events, all which might distort simple impact measurements and say little about whether grants focus on important problems (Bornmann, 2017).

Instead of focusing on differences in skill, talent and performance, perhaps we should ask: Does funding concentration lead to desirable or undesirable outcomes, and at what point do one outweigh the other? Concentration may be harmful on the basis of pure instrumental considerations, i.e. the effects of strong concentration render it counterproductive even if it is not "unfair". Concentration is perhaps not inherently bad, but may produce outcomes that are collectively undesirable. We could acknowledge that strong concentration of funding is a natural outcome of differences in skill, talent, effort, and a peer review-based selection of research projects, but still argue that a higher degree of dispersal would bring some beneficial effects outweighing the problems.

The results presented in this dissertation have some bearing on this debate. Analysing concentration from the perspective of research content and not only researchers, highlight some less obvious "epistemic" (Aagaard et al., 2020) effects of strong concentration and path dependency. While some degree of concentration likely is necessary to foster research talent and ensure originality in what is studied, the interdependence of individual and topical concentration, and the role of priorities in exacerbating these provide some counterarguments.

A stronger degree of funding concentration on individual researchers may serve to fund science with the greatest potential for publication, which should garner more citations. From the perspective of individual researchers' career development, this is important when publications and citations function as measures of recognition. However, a maximisation of individual recognition can be less ideal from a societal perspective, if it leads to less diversity in what is studied. This seems to be happening in the cases I have studied, and other empirical studies lend further support to this observation. Focusing on already well-studied chemical compounds or genes may lead to more success in funding competitions (Stoeger et al., 2018), but can perhaps be inefficient in making new discoveries or extraordinary innovations (Foster et al., 2015; Rzhetsky et al., 2015; Stoeger & Nunes Amaral, 2020).

Furthermore, it can be questioned whether the type of research emerging from current funding mechanisms is to the benefit of society (Bozeman, 2020). The results pertaining to funding of disease-related research in Article B directly addresses this. From a narrow perspective, a high concentration of funding on diseases such as diabetes or breast cancer makes sense because these diseases have posed a significant health problem in Denmark. But in the greater scheme of all societal health burdens, this concentration may relatively impoverish other research areas, e.g. back and neck pain or liver diseases brought on by alcohol and drug abuse. These are also diseases with a long-standing impact on population health. Because both private and public funders in Denmark seem to "underfund" research in to these diseases, the national funding portfolio appears unduly skewed. Such patterns turn the spotlight on who gets to define scientific topics worthy of funding, who exerts influence in adjusting budgets and the amounts allocated, and what type of research remains understudied or not studied at all. The term "undone science" have been used to describe the type of research that remains unfunded despite claims from societal organisations or movements of its credibility and worthiness (Frickel et al., 2010). Some science may necessarily remain unfunded or attract very little funding, exactly for the reasons described by Polanyi. Some topics may provide little opportunity for discovery, may not possess very high plausibility, or may provide little potential for original ideas. But we can probably think of topics, like neck pain and liver disease, which possesses value to societal actors such as patient groups and also remains a low priority in terms of funding.

Concentration of funding in an area with both strong demand and strong interests can have both negative and positive impacts on society depending on who benefits (Bozeman, 2020). External influence over funding portfolio becomes a study of power and resources in who has the authority to allocate funds and who can sway that authority. Here the study of funding concentration is also a question of the distribution of power, resources and opportunities that structure agenda setting in the scientific domain (Hess, 2009; Hess, 2015; Frickel et al., 2010, p. 4650).

Finally, the empirical analyses of this dissertation also shed light on more "unintended" potential consequences of concentration. In both Article B and D, we show that different funders have surprisingly similar priorities and tend to fund the same group of researchers and topics. These results highlight that uncoordinated decisions across funders, which may all focus on funding excellent and relevant science, can perhaps accelerate concentration to levels not desirable by any actor in the system. So while some degree of concentration can be desirable in individual funding competitions or funding institutions to weed out non-relevant or low quality research, this may multiply across the funding system and leave certain researchers with few external funding options or some topics and disciplines with a less than ideal base for building research capacity.

6.3 Making due with what we have

In the preface to this dissertation, I argued that discussions surrounding funding policy often lack a firm grounding in evidence. This is indeed puzzling as many of the contributors and participants in these discussions are researchers themselves. The development of funding policies of course needs to take the lived experiences of researchers into account, but it is my distinct impression that discussions often lack a broader perspective and that both policy makers, funders, and researchers would benefit from a more systematic approach to teasing out the ins and out of the research funding system.

While this project has had the explicit goal of providing an empirical perspective on some of the implications of funding science through competitive means, a central question is to what extent this goal has been met? Specifically, I would like to reflect on some of the shortcomings of the approach I have taken.

A large barrier to studying to what extent competitive research funding is concentrated across particular types of research content, as well as the underlying drivers of such concentration levels, are the lack of information on (a) what type of research was applied for, but not funded, and (b) what type of research was not applied for because of anticipation of the funding process?

Part (a) constitutes a problem for studying funding distributions because the level of funding (either number of grants or the value of grants) are conditional on a positive funding decision. What we want to estimate is the level of funding allocated to different topics, but the data I have gathered only permits us to observe the level of funding for topics *given* that they were funded. The problem becomes a lack of base rate in terms of how many grants were applied for on a given topic. For estimating how much funding is allocated to different topics, the fact that we only observe funded grants and not rejected applications constitutes a sample selection problem. While it clearly gives us an indication of the degree of concentration in funded research, it is not possible to pinpoint with certainty which part of the competitive funding process that creates concentration. Is concentration a function of sorting in peer review processes, where applications on some topics are less likely to pass review? Or is it a function of a lack of applications on certain topics? This sample selection problem is solvable. Data on rejected proposals would be helpful in determining the stage at which concentration is introduced[3]. It could help tease apart questions of the effects of peer review and unequal recognition. However, this is not a problem endemic to the study of research funding, but also applies to many studies of publication and citation dynamics. Studies of the diversity in what is published in terms of disciplines or topics are left with data on publications that actually passed peer review and was published in a journal.

Part (b) provides a similar problem, and would not even be mitigated by having access to data on unsuccessful proposals. There could potentially exist grant applications on various topics, which were never submitted, finished or simply remained an idea. It complicates any analysis on how researchers anticipate funding competitions and perhaps tailor their topic choices towards more "fundable" themes. Interviews with researchers tends to confirm that such anticipation takes place (Laudel, 2006b; Leisyte et al., 2010), but large scale investigations of such behaviours remain difficult. While having data on rejected proposal could provide evidence of "topical biases" in peer review as some recent research suggests (Hoppe et al., 2019)[4], it would not help in discerning structural determinants of funding concentration.

Another problem is the lack of suitable and comparable data across different contexts, both countries, funders, and research fields. Prior studies of funding concentration have had a clear focus on funding allocations in the U.S. National Institutes of

[3] I did, over a period of two years, attempt to get access to data on rejected or unfunded applications through dialog with different research councils, albeit unsuccessfully.

[4] This debate is not yet settled, as competing evidence shows that research on different topics do not significantly differ in review scores, only in funding propensities because of limited budgets (M. S. Lauer et al., 2021).

Health, because of available data on what is funded and metadata pertaining to grants and applications (Berg, 2012; Hand, 2012; Hand & Wadman, 2008; Katz & Matter, 2020; Lerchenmueller & Sorenson, 2018; Wahls, 2016; Wahls, 2018b). A clear objective of this dissertation has been to provide evidence beyond biomedical sciences from across research fields such as physical sciences, biological sciences, social science, and humanities. But in doing so, some context-dependent patterns and idiosyncrasies have probably been lost or glossed over. Furthermore, I have attempted to address questions of funding concentration outside the U.S. context by focusing on Denmark and the United Kingdom. Even if some patterns seem to reflect what have been found in other countries (e.g. Ma et al., 2015; Mongeon et al., 2016), the results presented here may be specific to the exact national research funding systems.

As noted in chapter 4, the Danish case is exceptional with its density of strong private research foundations with high investments in especially biomedical research, but also engineering and chemistry. This combination of private and public funders engaging in the sponsoring of research in universities is likely special to the Danish system. In the UK, private funders such as the Wellcome Trust or British Heart Foundation are major contributors to the total research budget as well, but it is not clear if they influence the direction of research in a similar way. Major research foundations in the UK are philanthropic or charity-based (e.g. Wellcome Trust, Leverhulme Trust, British Heart Foundation, and Cancer Research UK), while most of the Danish research funding foundations have explicit commercial interests through their ownership of research-heavy companies within pharmaceuticals, chemistry, or construction and engineering. While philanthropic foundations such as Wellcome Trust and Leverhulme Trust draw on finances from the pharmaceutical sector or via shares in multinational companies, they were not set up to control company shares. Similarly, while the British Heart Foundation explicitly prioritise medical research pertaining to heart conditions, neither of these funders are bound by market logics. The strong overlap in research funded by private and public research funders in Denmark could also be a product of a historically well-established research tradition within e.g. diabetes research, and the fact that many of the private research foundations only recently increased their share of funding to university researchers significantly.

Similar differences between the Danish and UK case exist for governmental prioritisation. As argued in chapter 4, where research councils in the UK have been kept in charge of implementing priorities through their traditional funding competitions, governments in Denmark delegated this task to particular strategic research councils, and have only recently allocated funding to earmarked topics within the Independent Research Fund Denmark. It is therefore not clear to what degree we can compare the effects of targeted or prioritised funding schemes in the two countries. Priorities enacted through traditional councils ensure that research councils act as important intermediaries between government strategy and researcher choices, and research councils could be more susceptible to researcher demands.

Lastly, the result presented here have relied on a small selection of methods for categorising the research content of individual research grants. As detailed extensively in chapter 2 and 4, these approaches are only a limited perspective on the aboutness of grants based on specific properties, which may not adequately describe all research content. Using the publications of grant holders as a way of characterising their research, exposes the categorisation to some of the important laws of bibliometrics, which tend to always produce core-scatter pattern in data. In an effort to improve the categorisation of grants, it would be beneficial to use many approaches to topic and speciality clustering in grants, and assess the degree of overlap and differences between them. It is likely that no method is ideal in all circumstances or that a combination of approaches could improve categorisation and labelling of grant topics (see e.g. Ahlgren et al., 2020; Waltman et al., 2020). Experimentation and evaluation of different categorisation schemes in collaboration with the grantees themselves could help facilitate a better portfolio overview for researchers and funders (Held et al., 2021).

6.4 Implications for research policy

While I have shown some clear patterns towards stronger funding concentration at the research content level, the question of how to handle these is a political one. Concentration in itself may not pose a problem for research policy, but here I have argued that concentration in certain research areas may be problematic and indicative of a need for policy revision. Besides investigating the degree of concentration, the aim of this dissertation has also been to question why concentration on certain topics and disciplines arise, and suggests some lessons for science policy-makers.

A first lesson centres on the number of available sources of competitive research funding. Contrary to what we might think, diversity in the funding landscape may not translate into diversity in what is funded. At least, this seems to be one driver of funding concentration in the Danish research system. When most funders value similar types of research or seek to support the same type of researchers, the result is likely to be a higher degree of concentration than perhaps any individual funder had intended. This unintended strong concentration is a product of policy choices (the use of certain shared standards of evaluation, a strict focus on excellence, and quality assessments including past achievements) with non-deliberate implications. A first factor is the relative conformity of the dominant research quality criteria employed by many funders when assessing research project proposals. When the majority of funding organisations base their funding decisions on a narrow set of criteria for research excellence, judged by elite peers and often through metrics of past performance (publications, citations, and grants), funding priorities will likely exhibit large degrees of overlap between different funders.

Concentration then ensues because multiple funding organisations aim to support the "top" or "best" researchers. This may be justifiable from the perspective of individual projects, but may become undesirable from a systemic point of view. A second supplementing factor may be the isolation of each individual funding decision. Funding organisations often choose between many competing projects of almost equal quality or review score, but have little knowledge of funding decisions outside their own organisation. Lack of oversight, transparency and knowledge of a project proposer's success in other funding bodies may thus amplify the concentration of funding for both individuals and topics. Researchers may have multiple project proposals evaluated at different funding organisations, or may attempt to spread risk by applying with slightly different projects at different funding sources. Funding each proposal can be warranted in isolation, but many funders may inadvertently end up supporting the same researchers, narrow topics, or even projects.

A second lesson revolves around the tendency for prioritised funding to mirror existing distributions of funding. The results presented here show some overlap at the systemic level, but also that individual researchers are difficult to move from their preferred topics. This dual stasis in both the overall overlap in distributions and individual focus can be counterproductive to the realisation of a more diverse topical and disciplinary funding distribution. It is not clear from the evidence how the disciplinary and topical targets emerge. In the case of the Danish strategic funding councils, priorities tend to be relatively broad and includes e.g. energy research, food science, transport and infrastructure, or disease and health research (Strategic Research Council, 2014). In the UK's Biotechnology and Biological Sciences Research Council, the priorities are more focused, and encompasses e.g. combatting antimicrobial resistance, integrative microbiome research, or reducing waste in the food chain (BBSRC, 2020). These are still not priorities at the topical level, but still targets more specific sub-disciplines. Broad and more vaguely defined priorities are often preferred by researchers because they leave room for researcher driven problem choice, but this may also have unintended consequences for funding concentration. When priorities are vague, they are easier to game for the already well-off, and existing research lines can be rebranded in order to obtain funding (Laudel, 2006b; Leisyte et al., 2010; Luukkonen & Thomas, 2016). This is one potential explanation of why priorities appear to overlap between strategic and non-strategic funding councils.

Countering tendencies for mirrored funding priorities is hard. Many proposed solutions fly in the face of traditional scientific norms of meritocratic allocation. One radical proposal has been to skip proposal competitions altogether, and fund people instead of projects (Ioannidis, 2011). In terms of funding concentration, a focus on funding researchers themselves can have either ameliorating or amplifying effects. Other proposals would institute modified lotteries, whereby proposals would be ranked by peer-reviewers, but funding would be awarded through a lottery among all proposals deemed of sufficient quality (Fang & Casadevall, 2016). Split-hair decisions among projects near the pay-line can often hinge on reviewer idiosyncrasies or assessment of the applicants past achievements, and some studies suggest that reviewers are bad at predicting which projects become more successful (Fang & Casadevall, 2016, pp. 3–4). However, the modified lottery does not ameliorate a key driver of funding concentration: writing the proposal itself. Thinking back to chapter 3, we may argue that incentives to study a narrow set of topics do not change based on the introduction of funding lotteries, when some peer-review is still in state. What researchers deem "fundable" may still direct the type of application and project proposal they write. Moreover, it does not reduce the burden and costs of writing applications in the first place. Researchers spend considerable time on applying for funding, encroaching their time allocated to doing actual research (K. Gross & Bergstrom, 2018). This may be less of problem for already well-funded researchers, who can strategically pick the funding competitions they deem worthy of such an investment, but it may further concentration by not lessening the burden for others. Another proposal has been to focus on excellence in peer review, but broaden the notions of what is considered quality. This could be a renewed focus on e.g. projects aimed at neglected topics of broad societal relevance. Traditional excellence orientation is likely to reward already successful researchers, assign even more funding to research topics that are already very well-supported, reinforce rigid disciplinary boundaries, and detach research from broader societal problems problems (Jones & Wilsdon, 2018, pp. 56–58; Madsen & Aagaard, 2020). If funders start operating with broader understandings of what constitutes research quality, it could help support more diverse research systems, with greater attention to societal expectations.

In general, funders could also eliminate potential deficiencies of the system if they adopted a research portfolio perspective on what they fund (Klavans & Boyack, 2017a; Wallace & Ràfols, 2015; Waltman et al., 2019). This includes deliberate post-funding analysis of what type of research was funded during the year, identifying possible gaps, and openly discussing these with researchers, other funders, and policymakers. Better overview of what type of research is funded could pave the way for more evidence-based priority setting, with a view to increase critical mass within important areas that have slipped through the cracks in the funding system. A strategic effort towards transparency in what and who is funded with both public and private means, and a greater diversity in funding schemes not only aimed at the "excellent" can be a crucial solution in future funding policy. It can be argued that better balance between dispersal and concentration will be advantageous from a scientific point of view with the added benefit of aligning the system with the needs of the people providing the resources to begin with.

Research articles

7. Diversity or Disparity? The Concentration of Funded Research Topics in the United Kingdom

Emil Bargmann Madsen[1]

The misalignment between societal needs and the priorities of conducted research have spurred calls for more attention to what types of research is actually funded. At the same time, published research exhibits strong path dependency by becoming more focused on well-established and reoccurring topics. The distribution of attention over various topics reflect the broader political economy and power structures in science, but whether path dependency starts already at the funding stage is largely unknown. In this article, I ask how skewed and path dependent the distribution of research funding is with respect to research topics. Using data on public funding of more than 30,000 research projects in the United Kingdom since 2006, I show that competitive public research funding is consistently concentrated on a minority of topics. These privileged topics continue to attract the majority of funding over time, and generally stay within the top parts of the funding distribution. Funding of research topics thus seems to follow a process of cumulative advantage, but while higher prior funding levels predicts higher subsequent funding, there is a decreasing marginal return.

[1]Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University

7.1 Introduction

The current system for allocating research funding is often described as facing a "prolonged crisis" (Fang & Casadevall, 2016) or "systematic flaws" (Alberts et al., 2014). The problems underlying these descriptions are many, but often include a growing concern over falling success rates and concentration of funding in the hands of the few (Fortin & Currie, 2013; K. Gross & Bergstrom, 2018; Peifer, 2017; Vaesen & Katzav, 2017; Wahls, 2018b). These concerns have largely been corroborated empirically. A wealth of research demonstrate cumulative advantage or "Matthew Effects", where narrow groups of researchers and research organisations accumulate disproportionate amounts of research funding. Funding distributions are extremely skewed (Katz & Matter, 2020; Larivière et al., 2010, p. 49; Ma et al., 2015, p. 14761; Mongeon et al., 2016; Peifer, 2017; Wahls, 2018b; Wahls, 2018a) and individuals in the top of the distribution tend to accumulate more funding compared to the bottom over time (Jacob & Lefgren, 2011; Katz & Matter, 2020; Lerchenmueller & Sorenson, 2018).

However, the focus on individual Matthew effects have neglected that inefficiencies in science funding may be most pressing at the research topic level. The distribution of funding to different research topics and problems is important because it indicates how a society values different types of scientific progress and research strategies (Klavans & Boyack, 2017b). Studies have documented wide-spread disparities between societal needs and research priorities (Evans et al., 2014; C. Gross et al., 1999; Manton et al., 2009; Yao et al., 2015; Yegros-Yegros et al., 2018). The biomedical sciences experience marked stratification in funding of disease-specific research (Evans et al., 2014; Head et al., 2013; Yao et al., 2015). Funding is also skewed towards translational biomedicine and pharmaceuticals, while neglecting public health and prevention studies (Head et al., 2016, p. 184; Jones & Wilsdon, 2018, pp. 15–18). This tendency is likely not endemic to the medical sciences. Similar concerns have been aired from researchers in physics, where directed funding to nanotechnology, energy materials, and photonics may crowd out funding for alternative basic research (Nature, 2003). Contrary to Matthew Effects for individual researchers, we are still somewhat amiss regarding the distribution of funding across research topics. The tendency for funding to concentrate in few research topics is often postulated (Alberts et al., 2014; Fortunato et al., 2018), but seldom put to the test. A recent exception, focusing on US funding 2008-2013, have nevertheless documented a considerably skewed distribution of funding over topics (Klavans & Boyack, 2017a). The concentration of funding correlate well with the prominence of topics in the scholarly literature but also the amount of funding a topic has previously attained (Klavans & Boyack, 2017a, pp. 1167–1169). Prominent topics accumulate funding while non-prominent topics are increasingly disadvantaged over time. However, quantitative assessments of funding distributions are often focused on more limited research disciplines (Head et al., 2013; Head et al., 2015; Hoonlor et al., 2013), the US or sub-national contexts (Best, 2012a; Bromham et al., 2016; Klavans & Boyack, 2017a; Manton et al., 2009; Mongeon et al., 2016; Sampat, 2012; Yao et al., 2015), and limited time periods (Klavans & Boyack, 2017a). I expand upon these studies by presenting data on 30,662 individual grants awarded by public research councils in the United Kingdom from 2006 until 2017. I aim to provide a broad description of *the extent that research funding is concentrated in research topics and how concentration have changed over* time by focusing on a diverse set of disciplines, and a much longer time period than previously considered.

I proceed in four steps. First, I summarise the empirical efforts to study funding concentration for individuals and argue for why we should start considering similar dynamics for research topics. Second, I outline a procedure for linking grants to topics through analysis of publications by the grantees. Third, I present evidence for a concentration of funding in topics over time, and how funding distributions become static. Lastly, I discuss some possible drivers of cumulative advantage for topics.

7.2 Empirical perspectives on funding and topic concentration

Examinations of how scholarly attention is distributed have largely followed two separate tracks. Firstly, researchers have focused on how certain individuals and institutions attract disproportionately large amounts of grants and resources. Secondly, a growing orthodoxy have been documented in different scientific disciplines where the same research questions are seemingly re-examined again and again at an increasing rate. Few studies have however joined these perspectives. This is important as these two processes are likely not independent. The concentration of funding in individuals and organisations likely privileges the topics studied by these. Conversely, a concentration in topics may give unequal funding opportunities for researchers acting in different fields and drive cumulative advantage at the individual level.

I start out by reviewing the empirical evidence of cumulative advantage in funding, pointing to three recurring patterns: Extreme skew in funding amounts and number of grants, stasis or path dependency of unequal funding distributions over time, and tendency for prestige to drive concentration. Next, I argue that these tendencies may carry over to concentration in topics funded.

7.2.1 Cumulative advantage in research funding

Early accounts from sociologists of science argued that competition for recognition between autonomous scientists would secure an efficient allocation of funds and efforts among scientific problems (Hagstrom, 1974; Merton, 1957; Polanyi, 1962). This view casts the scientific community as a self-regulating system effectively distributing attention across research problems (Sarewitz & Pielke, 2007, p. 7). Robert Merton's work on the Matthew Effect seems to break with such a view. In Merton's characterisation of the allocation of rewards in science, recognition accrues to those who provide original contributions to science (Merton, 1957, p. 639). However, even though such contributions receive accolades, peer-recognition is often distributed with a gradient (S. Cole & Cole, 1967, p. 382; Merton, 1968, p. 56). Well-established scholars tend to accumulate recognition through publication, awards, and grants to a much higher degree than less well-known researchers, irrespective of talent or the intrinsic merit of their contributions (Bol et al., 2018, p. 1; Goldstone, 1979, p. 385). This process is self-reinforcing as recognition signals competence to those who distribute scientific resources (Reskin, 1977, p. 493; Zuckerman, 1967, p. 396). While Merton himself did not undertake any formal analysis, his hypothesis of cumulative advantage for individuals and organisations (Merton, 1988, p. 606) became the focus of almost all later empirical investigations. Early studies focused on the concentration of publications and citations of individual researchers (Allison et al., 1982; Allison & Stewart, 1974; S. Cole & Cole, 1967; Long, 1978; Reskin, 1977), but recognition through grants have lately become a strong focus in the literature. Overall, studies document three common tendencies in competitive grant competitions across funding bodies and countries: an extreme skew of cumulative funding distributions, wide-spread stasis, or path dependency, in who receives funding, and a strong influence of prestige at both the institutional and individual level.

Extreme skew. Nearly all studies find a significantly skewed distribution of research funds, in terms of both number of grants and amount of funding, across individuals and research institutions. A large proportion of studies rely on funding data from the National Institutes of Health (NIH), and typically find that 20 % of funded researchers receive around 50 % of funding amounts, while 10 % of research institutions accumulate 80 % of funding (Katz & Matter, 2020; Peifer, 2017; Wahls, 2018a, 2018b). Outside the US and NIH, the skew appears even more extreme. In Quebec, around 20-25 % of researchers in natural sciences received 80 % of external funds (Larivière et al., 2010; Mongeon et al., 2016), while 10-15 % of researchers accumulated 80 % of funds in the biomedical sciences (Larivière et al., 2010, p. 49). Funding from the Engineering and Physical Sciences Research Council (EPSRC) in the UK appear to be exhibit the most skew. From 1985 to 2013, 8 % of Principal Investigators (PIs) received half of all funding, while 90 % of funding was awarded to just 20 % of universities (Ma et al., 2015, p. 14761).

Stasis. Besides a varying, but generally, large skew of grants and funding amounts, a number of studies also show a deepening of these inequalities over time. Researchers funded by the NIH generally accumulate much more funding that non-winners with similar grant review scores. Some estimates show that R01 grantees accumulated an average of $227,000 five years after their initial grant, and $648,000 after six to ten years, and early-career grant recipients also have much larger propensities for receiving a mid-career grant up to 8 years after an initial grant (Jacob & Lefgren, 2011, pp. 1173–1174; Lerchenmueller & Sorenson, 2018, p. 1012). Similar effects have also been observed for prestigious Veni grants in the Netherlands. Grant winners, with review scores similar to high-ranking non-winners, had accumulated an average of €40,000 more within 4 years of receiving a grant. After 8 years, this gap had widened to €180,000 (Bol et al., 2018, p. 3). Individual grant winners in the NIH also showed a considerable stasis when it came to winning further NIH grants. Grant winners entering an 8 year period in the top 10 % of the funding distribution also had a 58 % chance of staying in the top ranked percentile (Katz & Matter, 2020, p. 14). These individual feedback effects also seem to manifest themselves across the whole system. In the NIH, the Gini coefficient for both individuals rose from 0.4 in 1985 to 0.5 in 2010, and from 0.81 to 0.86 for research organisations (Katz & Matter, 2020, p. 14). For the EPSRC, Gini coefficients rose from below 0.5 to nearly 0.7 in the same period, with inequality between universities being consistently higher than between PIs (Ma et al., 2015, p. 14761).

Prestige of institutions and individuals. One persistent factor of stratification in studies of funding distributions have been the prestige assigned to individuals and organisations. Prestigious US universities are 1.7 times as likely to get funding and to get one application funded each year compared to less prestigious institutions. The amount of funding for investigators from prestigious universities were on average 2.4 times higher and the mean annual award size was 1.5 times higher (Wahls, 2018b, pp. 3–4). This prestige effect may also give funded PI's an edge in the publication system, as journals with higher impact factors tend to publish NIH funded investigators with higher median funding levels (Katz & Matter, 2020, p. 11). In general, project funding schemes seem to be subject to strong gatekeeper functions. Prestigious research institutions not only accumulate more funding, but also impact the amount of funding awarded to others. For the EPSRC, Ma et al. show (Ma et al., 2015) that UK universities form funding 'rich clubs' where "… affiliations with a high number of grants tended preferentially to collaborate among each other, forming tightly interconnected communities…" (Ma et al., 2015, p. 14762). Whether individual researchers serve similar gatekeeper roles is still questionable. In the Canadian Research Councils, researchers who are centrally placed in co-authorship networks also attract more funding (Ebadi & Schiffauerova, 2015).

7.2.2 Concentration of attention and funding in research topics

Despite the focus on cumulative advantage for individuals and institutions in the extant literature, the idea that topics also accumulate grants in a differentiated manner is not new. Early studies of Matthew Effects already speculated that cumulative advantage for individual scientists might carry over to the type of topics these considered. A researcher occupying a central location in a research area retain productivity advantages. Established researchers tend to work on synthesizing or integrating research topics, and are able to build up 'cultural capital' that can be transferred into resources (Allison & Stewart, 1974; Hargens et al., 1980, pp. 68–69). Whether this is actually the case is still not clear. Recent studies suggest that the extreme skew and stasis of funding also affects the distribution of funding to topics. For example, 50-75 % of projects funded by the NIH relate to 5 % of human protein-coding genes. The likelihood of becoming a PI is much lower for researchers studying the least studied genes compared to those who study the most studied (10 % vs. 20-25 %) (Stoeger et al., 2018, p. 8). Similarly, 86 % of articles indexed in MEDLINE investigated pairs of chemical substances previously investigated, while 14 % investigated completely new chemical entities or new relationships between known substances (Foster et al., 2015, p. 886). Researchers' reliance on competitive funding have grown in recent years, and in the same period these researchers have become more conservative in the topics they explore, with more repeated topic combinations (Rzhetsky et al., 2015, pp. 14570–14571).

However, the connection between funding concentration for individual researchers and for topics may also run counter. There may also be a feedback in funding success from topics to individuals. Career advancement in relation to grants have been speculated to demand a steady production of publications through incremental contributions to the same research agenda. Researchers rarely make radical shift to an entirely new set of research topics, perhaps because demonstrating continuity and cumulative scholarship is the best career path for a researcher (Jia et al., 2017, p. 1). Studies of this are still scarce. In computer science, a 10 % increase in the number of published papers in a given topic is followed by a 75 % probability of an increase in the number of NSF grants awarded on the same topic (Hoonlor et al., 2013, p. 80). So far, Klavans and Boyack (2017a) have provided the broadest examination of topic and funding concentrations. They examine how topic prominence in publications correlates with later funding by using data from 314,095 federal research grants from 2008-2014. Calculating text similarity between grant titles and abstracts with publications in the Scopus database, they assigned the most similar topic to each grant (Klavans & Boyack, 2017a, p. 1167). They find that topics in scholarly articles differ trough their prominence (i.e. how much the topics are cited and viewed) (Klavans & Boyack, 2017a, p. 1165). How intensely a topic is funded is then both a function of its historical funding levels, and the prominence of that topic. Funding to a topic correlates extremely well with funding levels in earlier time periods and the historical prominence of a topic, with 73 % of the variation in funding levels being explained by these two factors (Klavans & Boyack, 2017a, pp. 1168–1169).

Only 3.7% of topics with low prominence received funding, while the 136 topics with the highest prominence, had an average of 184 U.S. authors, who received an average of $91,870 each (Klavans & Boyack, 2017a, pp. 1168–1169). Topics are thus very unequal in the attention they receive, receive differentiated levels of research funding, and this inequality seems to persist over time. This seems to mirror the three central tendencies described above. Funding to topics is skewed, static, and if topic prominence in publications reflect prestige, then prestige affect the amount of funding a topic receives. Moreover, topics attracting the bulk of funding in a given period are perhaps the topics currently deemed 'important'. The volume of publications on a topic follow current research policy, where earmarked funding for topics such as nanotechnology or bio-fuels also spark more research in these areas (Domenico et al., 2016, p. 7).

I expand on these points by analysing the distribution and landscape of public funding in the UK through time. The purpose of my investigation is three-fold: (i) To attempt to expand the study of cumulative advantage in science beyond individual researchers and institutions by focusing on cognitive communities arranged around common research topics. (ii) To inductively describe the skewness and stasis in how funding is allocated to research topics, and how funding may cluster in already well-funded and prestigious topic areas. (iii) To discuss how policy factors external to science itself may contribute to funding concentrations.

7.3 Materials and methods

The analysis relies on a unique database of individual grants awarded to researchers in the United Kingdom covering, as a minimum, the last decade, and 97,861 grants. To assess whether funded topics reflect a process of cumulative advantage in scientific resources, I proceed in three successive steps. First, I combine automated and manual collection of grant data from seven public research councils, and across a wealth of different research fields. Second, I connect individual grants to a set of probable research topics by utilising fine-grained topic information from 180,100 scholarly publications linked to a grant. Third, I map the spread of funded research topics and discuss the potential mechanisms affecting this spread. It is important to note that my data only documents the distributional outcomes of successful grants. Because I rely on publicly available information on funded grant proposals, I do not assess how the rejection of grant applications play into the cumulative advantage processes of scientific topics. Furthermore, I cannot account for strategic choices to not apply for funding or alternative means of financing research.

7.3.1 Data

To construct the database on individual grants, I used a number of different data sources. For the Research Councils UK (RCUK), I obtained all available funded projects from each of the eight research councils subsumed under the RCUK. Project-level data were available in the RCUK Gateway to Research Database from 2005 to 2017. I consider only grants with a research focus, and dropped all grants aimed at students, intramural grants, skill development, etc. (See Figure S1 for the sample of grants used). Each funded project is assigned to one principal investigator. In the case of larger centre grant given for the establishment of a research centre of excellence, two choices were made. Centre grants directed towards a certain research organisation (i.e. the PI is a rector, department head, etc.) were left out of the analysis. This ensures that no grant is linked to publications from a PI who might not have been using the grant. For centre grants giving group-leaders or PIs a possibility to create a new research centre, I included the grant.

I matched each principal investigator to their publications using two data sources. First, I used the RCUK Gateway to Research to identify self-reported outcomes of grants, and narrowed the search to only journal articles and reviews. This yielded a gross sample of 305,247 publications directly linked to a specific grant. I then matched these self-reported publications to article-level information through the in-house version of Clarivate Analytics' Web of Science database hosted at the Centre for Science and Technology Studies at Leiden University. The Leiden database is an augmented version of the Web of Science with a thorough author-name disambiguation assigning individual cluster IDs to publications from the same author. This set of PI-linked publication yielded 180,100 documents (See Figure S7.1 for an overview of the final set of grants).

7.3.2 Connecting topics and grants

Science encompasses different communities with researchers intuitively orienting their work towards specific communities or subdivisions of science. These subdivisions could be denoted as fields, specialities, or disciplines (Small & Griffith, 1974, p. 17). Furthermore, some hierarchical partition of these communities is possible: There is broad research fields such as Physics, more specialised, subsumed, disciplines such as Astrophysics or Condensed Matter Physics, and then individual topics within these disciplines. I define a topic as a cognitive space materialised as a collection of documents that focus on a common intellectual interest, such as a research problem (Klavans & Boyack, 2017a, p. 1159), and researchers working on these problems. Scientific documents have often been categorised based on the journal in which they appear, but journals are broad in scope, and may cover many topics, or a topic may be spread across multiple journals.

Instead, a document level-classification is needed. Besides the journal-based 250 research areas classified in the Web of Science, the Leiden database offers more detailed article-level classifications through 858 meso-level topic clusters.

For each principal investigator, I recorded their distribution of research topics by their published scholarly articles. The topic clusters are based on direct citation links between articles, where each article constitutes a node in a network with direct citations between papers constituting an edge between a pair of nodes. A Smart Local Moving (SLM) algorithm (Waltman & Van Eck, 2013) is applied for community detection and clustering into topics. Individual article nodes are moved into a partition and then aggregated so that each topic constitutes a node in an aggregated network. The full procedure is outlined in Waltman and van Eck (2012) and Traag et al. (2019). Their method outlines a procedure for constructing a set of micro-level topics or fields, and how these topics can then be aggregated to form a smaller set of meso-level or macro-level topics.

An example of the use of micro-level topics is the CWTS Leiden rank of universities where 4535 topics are used. I, instead, rely on the more aggregated 858 meso-level topics. Because each PI may work on a set of topics, or a grant relates to different areas, I calculated a topic weight for each grant with the most frequent topic in the publication set weighing the most. Using this weight, I split the grant amount and number of grants so to assign the weighted amount of money and grant numbers to a topic by year. The weighing of funding per grant can be illustrated through a simple, fictive example: A grant may be awarded for £100,000 and result in three publications within two different topics, which we denote topic A and topic B. If two of the publications are on topic A, and one is on topic B, the grant and the grant amount is weighted according to the frequency of these topics. Based on this one grant, topic A achieves 0.66 grants and £66,666.66, while topic B receives 0.33 grants and £33,333.33. I then sum up these weighted number of grants and amounts for each topic. This approach differs from other studies, where grant abstracts are compared to publication abstracts, and the most similar abstract topic is assigned to a grant (Klavans & Boyack, 2017a). Weighting grants according to multiple topics instead capture either the mono- or interdisciplinary character of a grant, and have previously been used for analysing topic diversity in the National Science Foundation (Nichols, 2014).

Figure 7.1 provides a topical map of the Research Council funding landscape, based on the outlined weighting method. Each bubble represents one of 858 topics, sized by the weighted amount of funding for each topic in the period 2005 to 2017. The distance between each topic indicates the relatedness of these based on citation links. Even though the algorithmically constructed research topics may ameliorate some problems with journal-based classification, they do carry other problems. For one, the Web of Science coverage of publications from different research areas are still skewed towards the

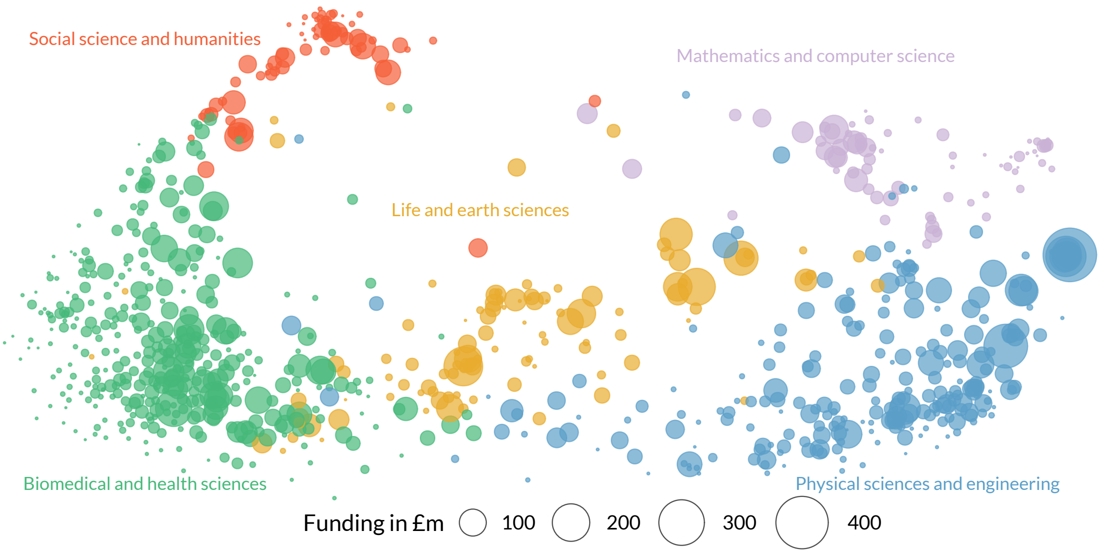


Figure 7.1: Map of funded topics. Funding amounts for 858 topics between 2005 and 2017

medical- and natural sciences. This may bias what type of grants I am able to link with a publication and topic. Second, the size of research topics in terms of number of publications differ. Observations of skewed funding distributions may then be a result of size differences in the publication input. However, these skewed distributions of papers and funding could also be endogenous to each other. Certain topics may be popular, resulting in both more publications and funding. To try to ensure that my results are robust to these problems, I recreate many of the main results for individual research areas by performing analyses on each research council.

7.3.3 Analytical approach

To explore how public research funding has been spread across research topics I conduct two sets of analyses. First, I focus on how the distribution of funding have unfolded with respect to the topics actually funded. This part neglects the span of topics not covered by funding from the national research councils but focuses on the inequality in topics actually funded. Given a lack of data on the research topics of grant applications not funded, or the topics focus of researchers who did not apply for funding, the analysis likely underestimate the funding inequality between topics. This is essentially a problem of self-selection. Topics with low funding success is not represented and researchers who anticipated low success because of their research topics are not accounted for.

While previous studies have examined inequality at the individual and institutional level (Katz & Matter, 2020; Mongeon et al., 2016; Wahls, 2018a), I focus on the topic level. Just as investigators participate in a funding 'economy' where grants are the 'income' (Katz & Matter, 2020), topics may similarly 'compete' for attention through funding. I calculate the cumulative distribution of funding amounts and the number of grants for all topics researched by funded scientists for the period 2006-2017. Furthermore, I investigate how the cumulative funding and grants distributions have changed over time. To assess the skewness of these distributions, I calculate three indices of inequality or dispersion of both number of grants and grant amounts. First, I employ one of the most popular indices for measuring income concentration; the Gini coefficient. The Gini coefficient have found widely applied in most studies of inequality in research funding (Katz & Matter, 2020; Ma et al., 2015) because of its direct relationship with the distributional Lorenz curve: It traces the ratio of the area between a perfect egalitarian Lorenz curve and the empirical curve to the total area under the egalitarian curve. However, the Gini is often criticised for being insufficiently sensitive to changes in the tails of a distribution (Atkinson et al., 2011, p. 10). In studies of income inequality, this has led to a greater focus on the share of income in different percentiles of a population, and the ratio of these shares. Here, differences between the very top and bottom have attracted particular attention. I therefore calculate funding and grant shares for percentiles of funded topics, and compute the Palma index of top 10 % funding shares vs. bottom 40 % funding shares (Cobham et al., 2016).

To investigate the cumulative advantage of topics over time, I attempt to assess the funding trajectory of each topic from the first year it was funded. The year a topic enters the pool of funded topics, it will enter a given percentile rank according the amount of funding allocated to the topic. I then calculate the percentile rank for a topic in all subsequent years, and show how topics that entered the distribution in different rank brackets tend to stay within those brackets.

Finally, I quantify how the funding of topics may exhibit path dependency and positive feedback by estimating a Bayesian hierarchical model. First off, this is complicated by how we measure "success" for a topic across the years of the funding data. In a given year, a topic can attract funding or not, which is in itself a measure of how successful the topic is. Given that it does attract funding, the size of the amount indicates something about the priority placed on that topic in the funding system. These two processes are essentially two linked events that together define the topic priorities in a year. However, a topic can attract zero funding both because no priority is placed on it, but a small amount of funding can also signify low priority. To accommodate this, I treat funding for topics as arising from multiple processes. Priority can be set both through no, or low levels of funding. I incorporate this in a model of path dependency by modelling the amount of funding to a topic as a mixture between a Bernoulli and Gamma process (a hurdle gamma model), i.e. a mixture of zero and positive values. I use past funding amounts for a topic to model how static the distribution of funding is. Because early establishment of prestige for some topics may affect the amount of funding for the topic later, I include a series of lagged levels of funding amounts for each topic for a given year. These lags measure the funding amounts for a topic i in 1- 5 years prior. The model is estimated as a hierarchical model with a random intercept for each individual topic. Including a varying intercept for each of the 858 topics entails at least two benefits. The random intercept allows me to quantify the within-topic path dependency in funding, by returning the average within-topic effect in a given year of different funding levels in 1-5 years prior. It takes into account that the amount of funding to a topic is measured repeatedly. Moreover, it gives an estimate of the variation in funding path dependency for all topics. The model is specified with the stochastic component

$$F_{it} \sim HGamma(\pi_{it}, \mu_{it}, \theta_i) \tag{7.1}$$

Where F_{it} denotes the amount of funding, for a topic i in year t. π_{it} is the probability of zero funding for a topic in a given year, and μ_{it}, θ_{it} describes the distribution of non-zero funding. Then the systematic component can be given as

$$F_{it} = \begin{cases} \pi_{it}, & \text{if } F_{it} = 0 \\ (1 - \pi_{it})Gamma(\mu_{it}, \theta_i), & \text{if } F_{it} > 0 \end{cases} \tag{7.2}$$

$$log\frac{\pi_{it}}{1 - \pi_{it}} = \beta_{0it} + \lambda_i + \sum_{k=1}^{5} \beta_k F_{ijt-k} \tag{7.3}$$

$$log\mu_{it} = \beta_{0it} + \lambda_i + \sum_{k=1}^{5} \beta_k F_{ijt-k} \tag{7.4}$$

With F_{ijt-k} being a vector of lagged funding levels such that $k = 1, 2, 3, 4, 5$, β_k are the fixed population-averaged/fixed effects of previous funding, and λ_i is the random intercept. All five variables of lagged funding amounts were mean-normalised, divided by one standard deviation, and were assigned a Cauchy prior probability such that

$$\begin{aligned} \beta_{0it} &\sim Cauchy(0, 20) \quad \text{[intercept]} \\ \beta_k &\sim Cauchy(0, 2.5) \quad \text{[lagged funding level]} \end{aligned}$$

In the supplementary material (figure S7.6), I report 12 models with alternative prior specifications. These models produce similar results. All posterior distributions were estimated empirically by Hamiltonian Markov Chain Monte Carlo (HMC) using R (ver. 3.6.2) and the package 'brms' (Bürkner, 2017, 2018). For all models, I assigned a 2,000 burn-in and 8,000 iterations for the sampler, using two chains, an acceptance rate of 0.9 (adapt delta), and a max tree depth of 20. The visualisations of the descriptives and posterior distributions were made with the R packages 'tidybayes' (Kay, 2019) and 'ggplot2' (Wickham, 2016).

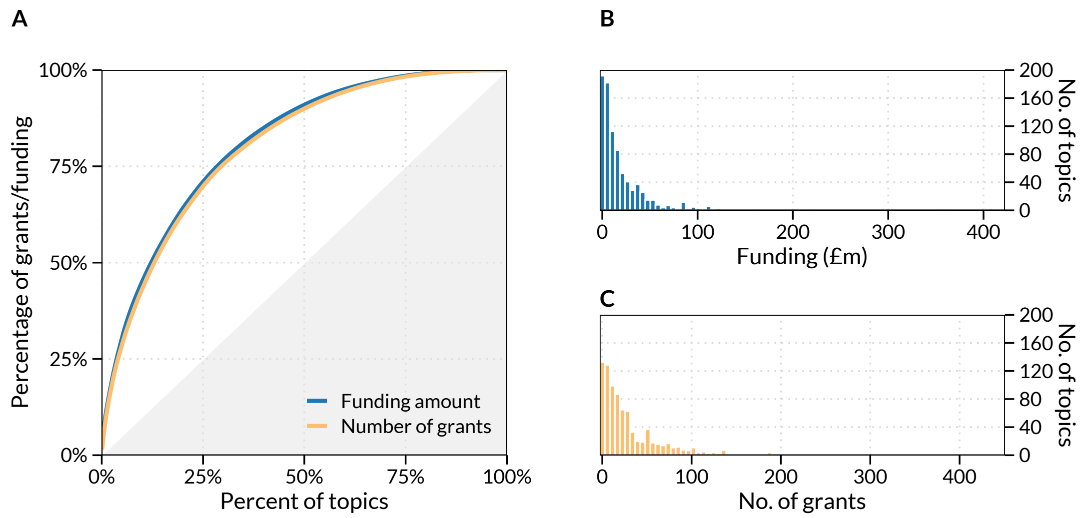


Figure 7.2: Distribution of funding to topics, 2005-2017. (A) Cumulative distribution of funding and grants to 858 topics. (B) Distribution of funding amounts. (C) Distribution of number of grants.

7.4 Results

I present the results of the analysis in two consecutive sections. The first section focus on the issue of skewness and concentration of funding to research topics. I present a range of descriptive evidence for the full time period (2006-2017) and show how concentration has been stable over time. The second section focus on the dynamic interpretation of cumulative advantage, and presents both descriptive and model-based evidence of stasis or path dependency.

7.4.1 The skewness of topic prominence in research funding

In order to investigate the concentration or dispersal across different research topics, I calculated the cumulative funding to each of the 858 topics across the entire period. Figure 7.2 shows the distributions of both the cumulative grant amounts, and the cumulative number of grants per topic. Figure 7.2A plots the cumulative distributions across all topics through a Lorenz curve. Complete equality of funding and grants would entail the two distributions tracking the 45-degree diagonal. Instead, I find a high degree of concentration with 10 % of topics amassing 44.8 % of funding and 43 % of grants, or 25 % of topics receiving 70 % of both funding amounts and grant numbers. 7.2B and 7.2C shows the two distributions of both funding amounts and grant numbers in absolute terms. In the supplementary material, Figure S7.2 shows Lorenz curves for each of the UK research councils included here. The concentration of funds in each council is even

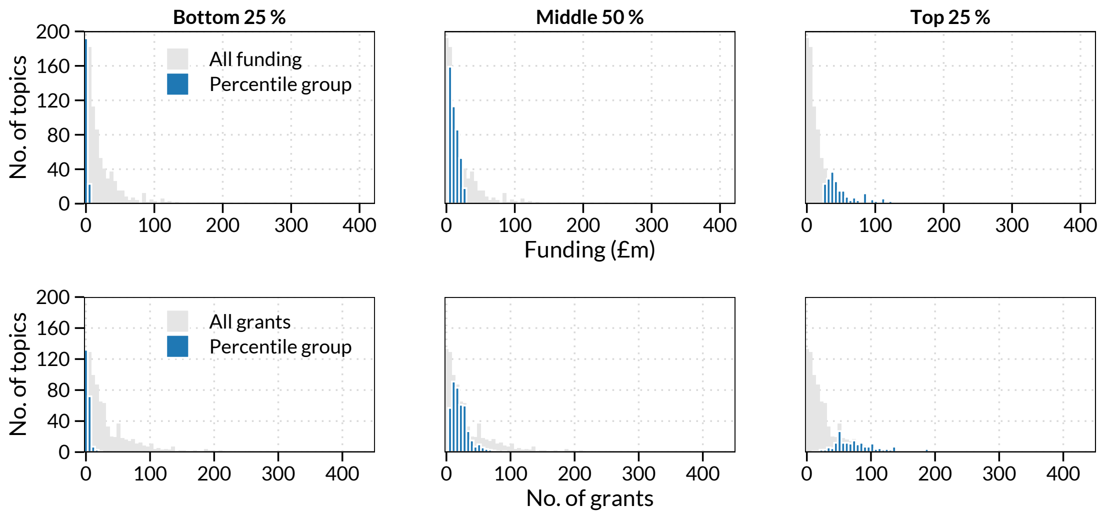


Figure 7.3: Distribution of funding and grants in percentile groups. Percentile groups are based on the distribution of funding amounts.

more skewed than across all councils. Consequently, the skewness of funding does not seem to be an artifact of the grant sample I am using.

Across the more than 30,000 grants, the research councils distributed close to £18 billion during the entire period, and 85 of the 858 research topics accumulated around £8.69 billion of the entire sum, while 214 topics received almost £12.8 billion. One of the driving factors behind the concentration of funds in a small number of topics may be the differences in monetary demands across fields of research. Derek de Solla Price first noted that the cost of conducting research had generally increased rapidly following the Second World War (Price, 1963). However, these costs are likely not evenly spread out. Some research questions are more expensive to answer than others. In the case of the distribution of funding, this might entail that the most well-off topics do not necessarily accumulate more grants, but just larger ones. Figure 7.3 shows that this is not the case. The top 25 % of funded topics are funded through a combination of many smaller grants and a few bigger grants.

Besides the tendency for high degrees of funding concentration, a related concern has been that changing research funding systems increasingly facilitate higher degrees of concentration. Generally, analyses have corroborated these concerns with respect to concentration of funds to individual researchers or institutions (Katz & Matter, 2020; Ma et al., 2015). To investigate a possible increasing concentration, Figure 7.4 plots three indicators of concentration across time. Figure 4.A shows that the proportion of funding awarded to the top 25 % of topics is actually more concentrated on a year-by-year basis than for the entire period.

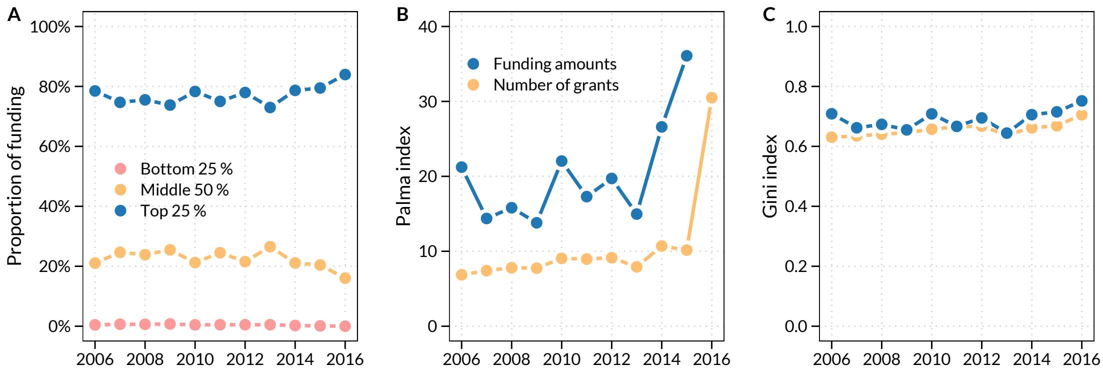


Figure 7.4: Concentration of funding and grants, 2006-2016. (A) Proportion of funding amounts in three percentile groups. (B) Palma index of funding amounts and grant numbers. (C) Gini index of funding amounts and grant numbers.

The top group of topics consistently attract close to 80 % of all funding, while the middle 50 % accounts for close to 20 %. The skewness of the yearly funding distributions are also reflected in the two inequality measures. The Palma index in Figure 7.4B show some fluctuation, but the top 10 % of topics attract, on average, 20 times the amount of funding awarded to the bottom 40 %. Similarly, the Gini coefficient in Figure 7.4C shows substantial concentration with a value well above 0.6. In sum, I find a stark concentration of research funding in a few research topics. While this may indicate at least some cumulative advantage for certain topics, it does not imply that the same research topics continuously accumulate more funding because of early success or path dependency. To answer this question more precisely, we need to look into the funding accumulation within topics, and not just across them. This latter point is the focus on the following section.

7.4.2 4.1 Stasis and path dependency for funded topics

The question of how much, if at all, individual topics accumulate funding unevenly over time is important in shedding light on cumulative advantage in research funding. This is essentially the central mechanism of Merton's Matthew Effect. Recognition is not just distributed in a skewed fashion, but also deepens as time goes by. It is also a central finding in much of the literature on topic concentration in research literature (Foster et al., 2015; Rzhetsky et al., 2015; Stoeger et al., 2018). Topics that proved important early are more likely to be funded and investigated later. To shed light on this mechanism, I make use of the time dynamics in the dataset of grants. In Figure 7.5, I calculate the funding percentile of each topic when first funded in the period 2006-2017. I group each topic into four percentile bins (Top 25 %, 50-75 %, 25-50 %, and Bottom 25 %) of 214- 215 topics each, and calculate each topic's funding percentile in the following five years.

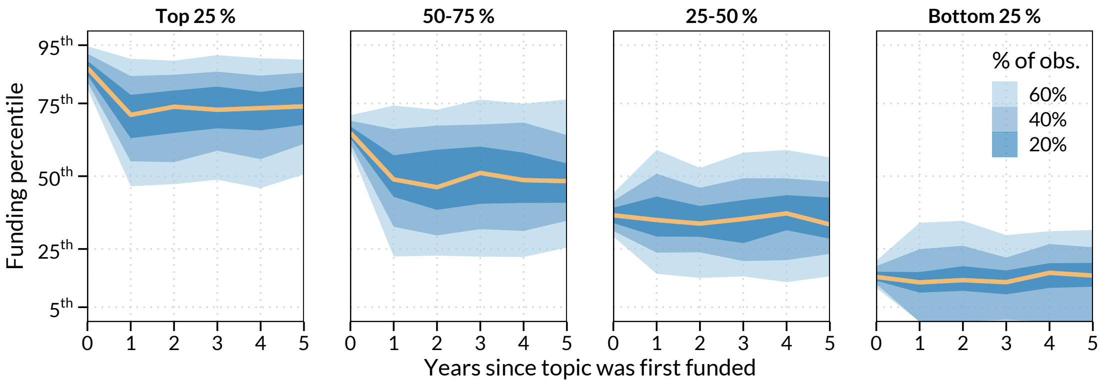


Figure 7.5: Subsequent funding percentiles for topics. Groups correspond to 25 % increments of the funding distribution when topics were first funded.

The central tendency in the figure is the median funding rank of each of the four groups, while the shaded areas include 20 %, 40 %, and 60 % of the percentile distribution. The topics entering the funding distribution within the top 25 % have a tendency to stay within the top in subsequent years. The median funding rank of this group is the 71st percentile one year after, and increases slightly to the 73rd percentile in each of the subsequent four years. Conversely, the group of topics starting out in the bottom 25 % tend to hover around the 15th percentile in the following five years. Figure S7.4 in the supplementary material repeats this analysis for groups of 10 % (with 85 topics in each), where the top 10 % are generally within the 85th funding percentile in the years after first being funded. Figure S5 shows that, within funding councils, the stasis of funding distributions is also evident within the biosciences (BBSRC), engineering and physics (EPSRC), and medical sciences (MRC); the three areas most likely to be covered in the Web of Science. Figure 7.6 shows the relative proportion of funding for each of the four groups each year after first funded. The top 25 %-group receive 66-71 % of the total funding amounts in the five years succeeding, while the 50-75 % group receive 22-25 % of the yearly funding. These results at least indicate some degree of path dependency and stasis in what research topics are prioritised in the competition for funding. The top group of topics tend to remain within the top of the funding distribution, while groups of topics with lower initial funding tend to remain in the lower percentiles of later funding distributions. However, Figure 7.5 also reveals a rather wide spread in what funding percentile the well-funded topics fall in to in the years after first entering the dataset. For the top 25 % of topics, 60 % span the 45-90th percentile interval of funding in the years after. For the top 10 % this interval is approximately the 78-89th percentiles. The spread of subsequent funding percentiles is less pronounced in the bottom of the distribution. 60 % of the topics in the bottom 10 % of initial funding, never rise to more than the first

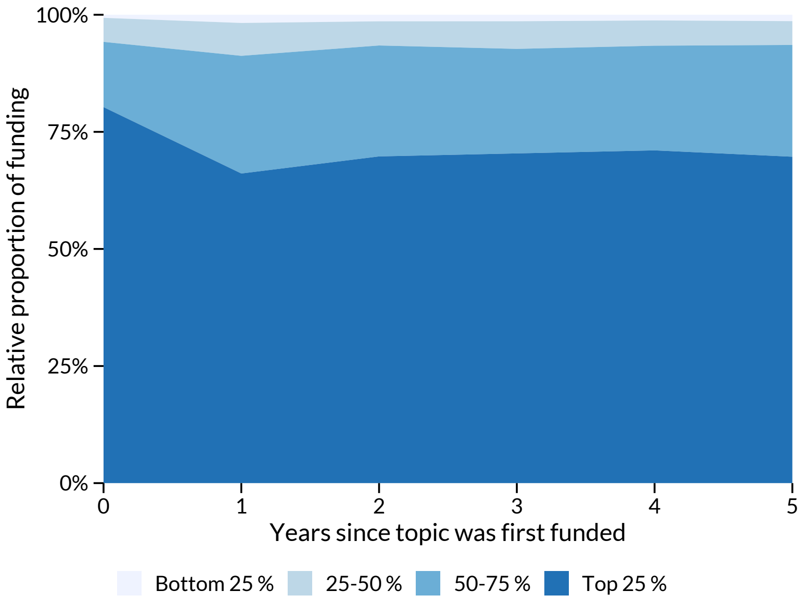


Figure 7.6: Relative proportion of funding for topics. Groups correspond to different parts of the funding distribution when topics were first funded.

25th funding percentile. This is not to say that topics with low initial funding never become more well-off. Some topics in the bottom 10 % actually end up around the 45th funding percentile.

My results seem to affirm that some cumulative advantages for topics do take place in the competition for funding, but it is by no mean a deterministic fact that the rich always get richer, while the poor stay poor. However, the topics in the absolute bottom or top of the distribution tend to stay within the bottom- or top half of later grant distributions. These tendencies are factors of different groups of topics. To zoom in on how individual topics accumulate funding, I end by presenting results from the statistical model of path dependency presented above. The model relates the amount of funding a topic receives to the past levels of funding for that topic. Figure 7.7 plots the population-average ("fixed" effects) amount of funding for a topic, based on the amount of funding in the previous one to five years. Each of the five panels show the expected mean amount of funding at different levels of previous funding. I show the expected mean funding in a given year for 0 to 20 million pounds of in previous years, as this approximately incorporates five standard deviations of previous funding (for all time lags). Based on increasing amounts of previous funding in one to four years, the model indicates a slight tendency for cumulative advantages. A level of previous funding of about five million pounds in one to four years prior approximately translates into an average of around 1 million pounds of funding in a given year. Based on the model, this result seem to be quite robust. Figure 7.7 shows the population average, with 95 % credible intervals, meaning that a premium of five million pounds of prior funding is highly likely to fall

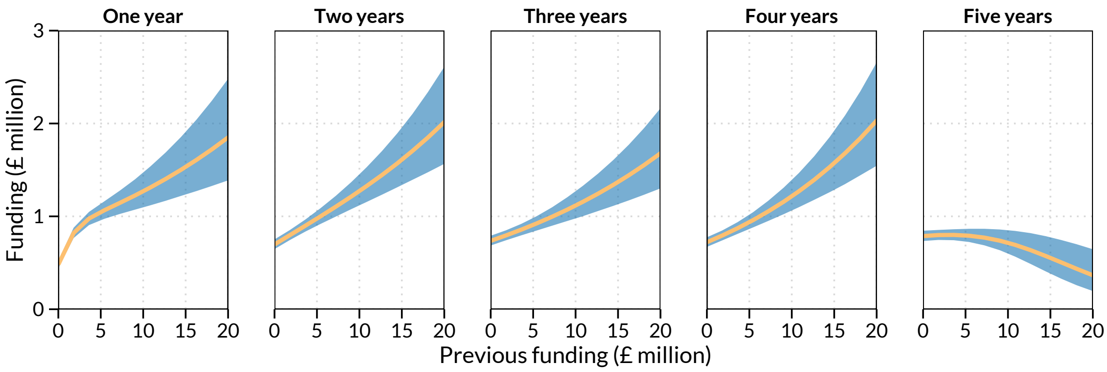


Figure 7.7: Average funding amounts for topics conditional on previous funding amounts. Each panel shows the population average posterior fit for previous funding one to five years before. Posterior fits based on model 3 in Table S7.2 with 95 % credible intervals.

within a range of £900,000 to £1,600,000. Another noteworthy finding is the apparent curvelinear relationship between current and past funding.

One example is the relationship between funding amounts now and two years prior. Here, the difference between no previous funding and five millions pounds of previous funding amounts to a difference in current funding around £400,000. Looking at the difference between £15 million and £20 million in previous funding, my model indicates a difference in current funding of almost £600,000. This increasing marginal return to previous funding is more pronounced for higher levels. Figure S7.7 is identical to figure 7.7, but with the average current funding estimated for previous funding levels of 50 to 90 million pounds. A difference in previous funding (two years prior) between 50 and 55 million pounds corresponds to a difference of around two million pounds. However, as the amount of previous funding increases so does the width of the credible intervals. For previous funding of 50 million pounds two years before, the premium could be as low as four million or as high as 15 million pounds.

Lastly, we should note the different effect funding five years prior seem to have on current funding levels. Instead of an increasing return to previous funding, past funding levels seem to matter little to current funding, or even be negatively related. For lower levels of previous funding (0-5 million pounds) a change from 0 to 5.4 million only increases current funding by £3,280, while a 5.4 million pound difference in previous funding at 21 million vs. 16 million translates into a decrease in current funding by £196,114.

7.5 Discussion and conclusion

The impact of competitively awarded research funding on the direction and content of science and the science system have attracted much attention. One persistent fear have been that researchers' increased reliance on competitive grant money leads to a narrowing of topics studied. Yet, we actually have little empirical evidence on how funding is distributed across different research topics and whether funding distributions are a prone to a cumulative advantage, which may exacerbate this narrowing. Here, I have tried to ascertain whether some research topics actually experience a form of cumulative advantage in funding, but I would argue that a definite answer is not yet warranted. Heeding the cautious words of earlier investigations of cumulative advantage in science: "Although the basic idea of cumulative advantage is quite simple, it is actually a rather complex hypothesis involving numerous causal links organized into several feedback loops. As a consequence, it is difficult to specify just what empirical findings would disconfirm the hypothesis" (Allison & Stewart, 1974, p. 622). My analyses do however indicate that some of the proposed components of a Matthew effect in funding of topics are present. Firstly, the distribution of funding to different research areas are highly skewed with 25 % of research topics amassing 70 % of all funding. This concentration are not simply due to a few topics being awarded a few large grants, it recurs across very different research fields, and is stable throughout the period. Secondly, stasis or path dependency is a central tendency in the distribution of funding across time, but with some exceptions. The top 10 % and top 25 % of topics in terms of funding tend to stay within the 85th and 75th funding percentile, while the topics in the bottom 25 % of the distribution generally remain in the bottom group. However, topics just outside of the top 25 % do occasionally end up in the top of the distribution, so path dependency seems to primarily affect the very top or bottom. Lastly, there is a slight premium and increasing return to funding for a topic based on previous funding. This is congruent with earlier findings that also show past funding levels to be the best predictor of future funding levels (Klavans & Boyack, 2017a). In this paper, I show that this relationship may be one of increasing marginal returns, with small amounts of prior funding leading to very small future funding levels, while very large amounts of prior funding yielding much large future returns. However, the amount of funding in prior and current time periods are not 1:1, as the predicted amount of funding is substantially lower than the previous levels. Figure S7.8 shows the difference between prior funding and predicted funding for all time lags. For one to four year lags of funding, previous funding needs to surpass 100 million pounds before the expected current funding amounts are larger.

Why some research topics seem to garner more attention than others is likely a question with multiple answers, but the results at least indicate that funding allocations are one piece of the puzzle. However, the analyses I have presented cannot discern the driving factors of cumulative advantage in science funding. I want to highlight two possible types of explanations. One is the different drivers of cumulative advantage that are built-in to scientific systems. These can broadly be thought of as the prestige effects I alluded to above, which have been the main focus of earlier investigations of funding concentration. Success in academia is often assed through past achievements, such as publications, grants, or awards. Many different studies show a concentration of funds on individual researchers or institutions, and document how certain researchers and research organisations continue to accumulate the bulk of all funding. This concentration is likely to spill-over into the type of research conducted. If few researchers and organisations receive the majority of funding, and also research a narrow set of topics, the concentration of funding in one domain may impact the other. Likewise, the past success of a research topic in the form of early breakthroughs, a well-developed research literature, and proved applicability, or many citations, may spur further funding of such topics because they appear promising.

Another type of drivers are systemic and caused by external pressures. This could be funders focusing on a more limited portfolio of topics they hope will lead to immediate inventions or solutions to societal problems. This is often argued to be the case in biomedicine, where translational research with close links to medical practice is prioritised over more fundamental research (Alberts et al., 2014, p. 5774). It may also be that researchers foresee what topics might be deemed "fundable" and self-select into research avenues they are more likely to obtain funding for (Gläser & Laudel, 2016; Laudel & Gläser, 2014). More worrying, topic selection and funder priorities may suppress specific topics as some researchers experience less funding success with novel research (Wang et al., 2018). Evidence from the NIH also suggests that reviewers favour certain topics even if topics with lower success are just as or more cited than high-success topics (Hoppe et al., 2019).

The results presented here could be used as an opportunity to discuss how research funding systems should be organised. Complete parity of funding for all topics is not, and should not, be a goal in itself, but the persistent skew and stasis in what topics are funded should indicate that some measures to ensure more diversity is warranted.

7.6 Supplementary material

7.6.1 Sample of grants used in analysis

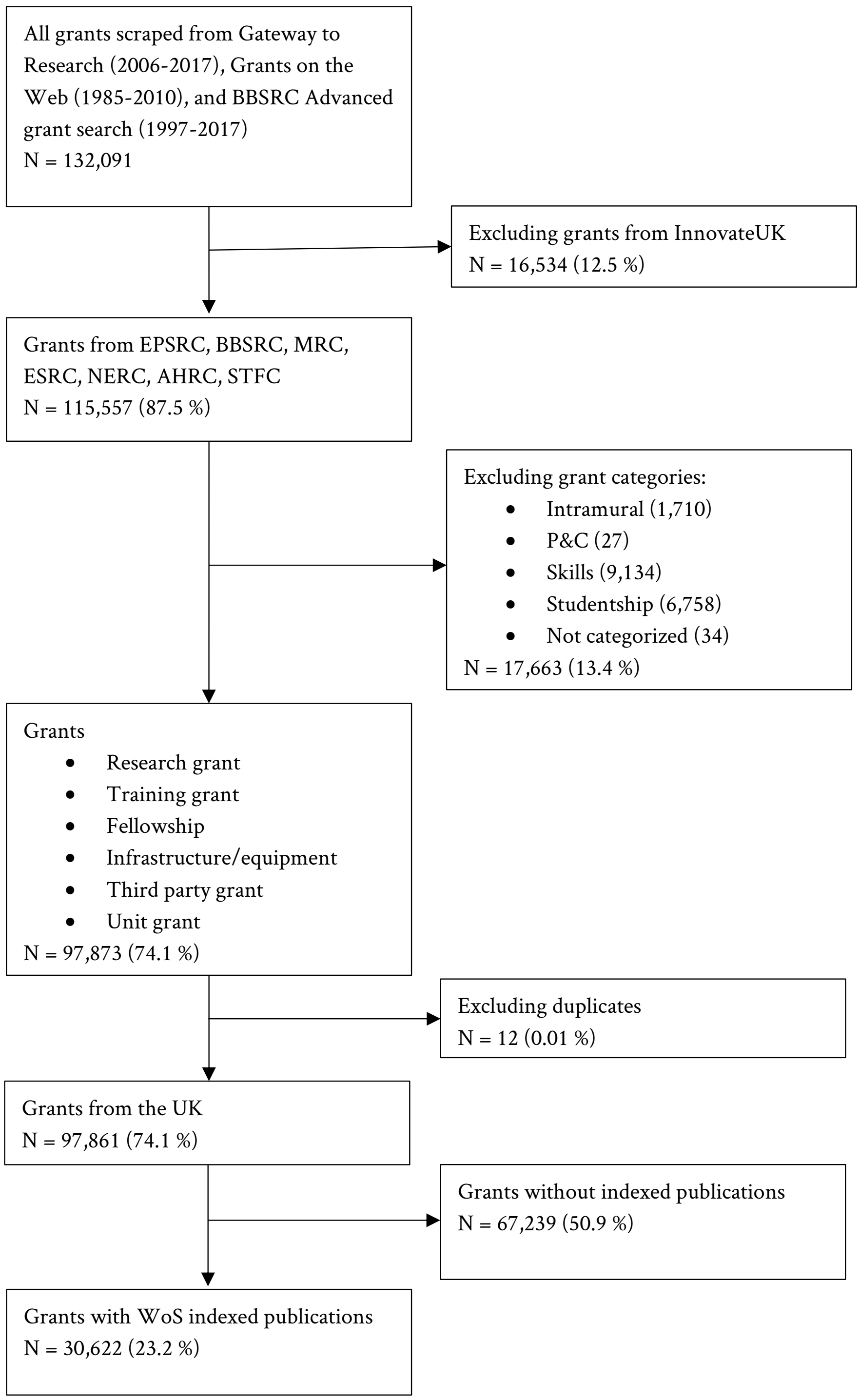


Figure S7.1: Final grant sample

7.6.2 Additional analyses of skew

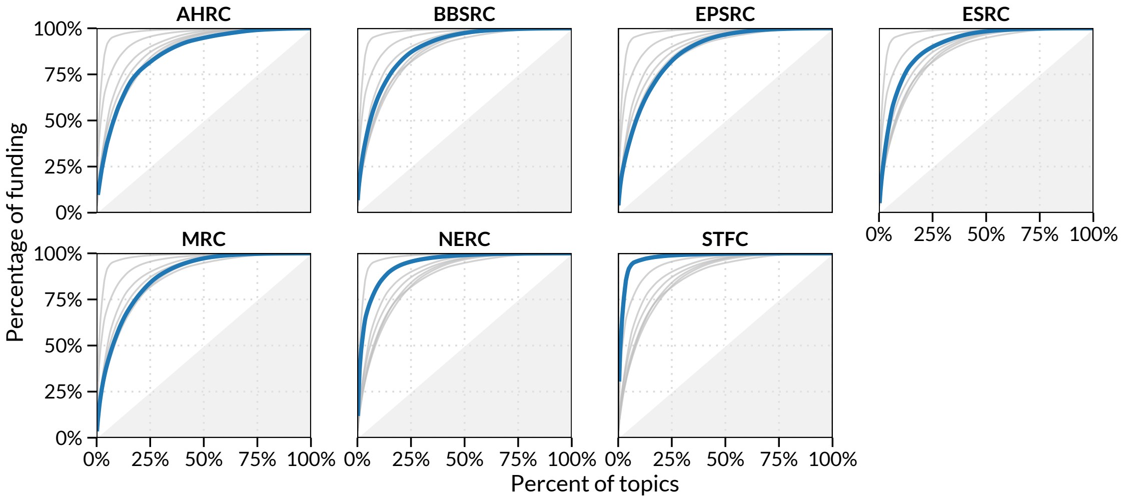


Figure S7.2: Cumulative distribution of funding amounts over topics for each research council.

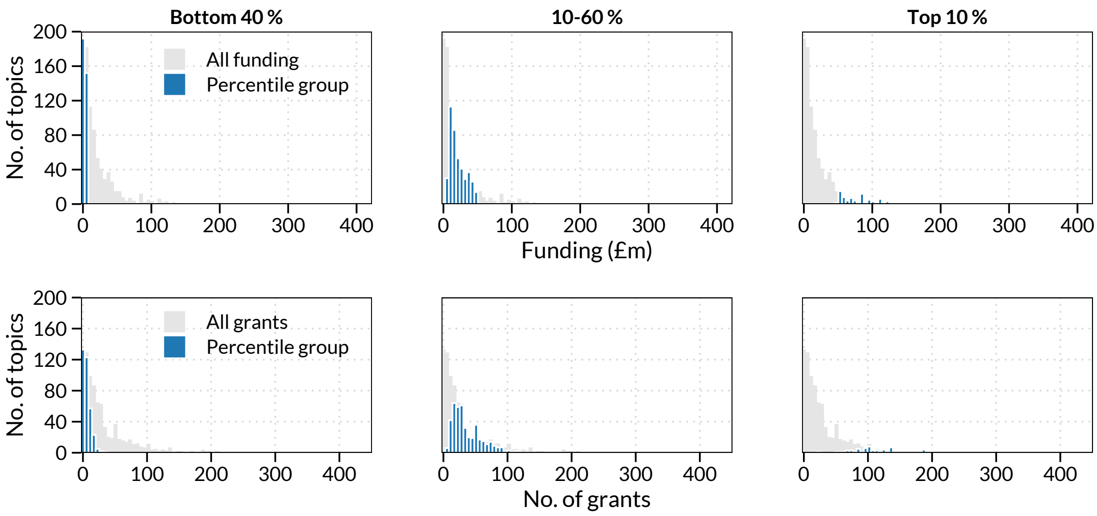


Figure S7.3: Distribution of funding and grants in percentile groups. Percentile groups are based on the distribution of funding amounts.

7.6.3 Additional analyses of stasis

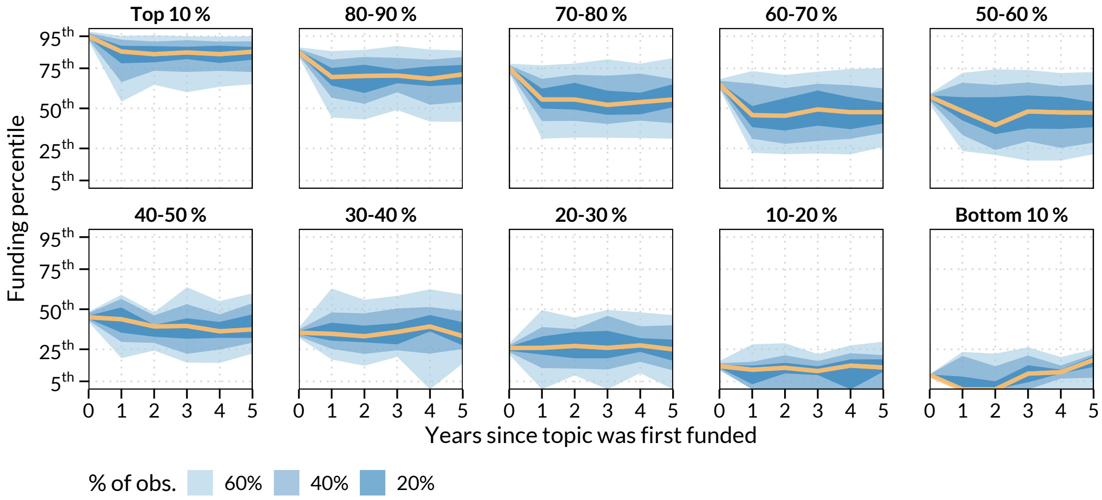


Figure S7.4: Subsequent funding percentiles for topics. Groups correspond to 10 % increments of the funding distribution when topics were first funded.

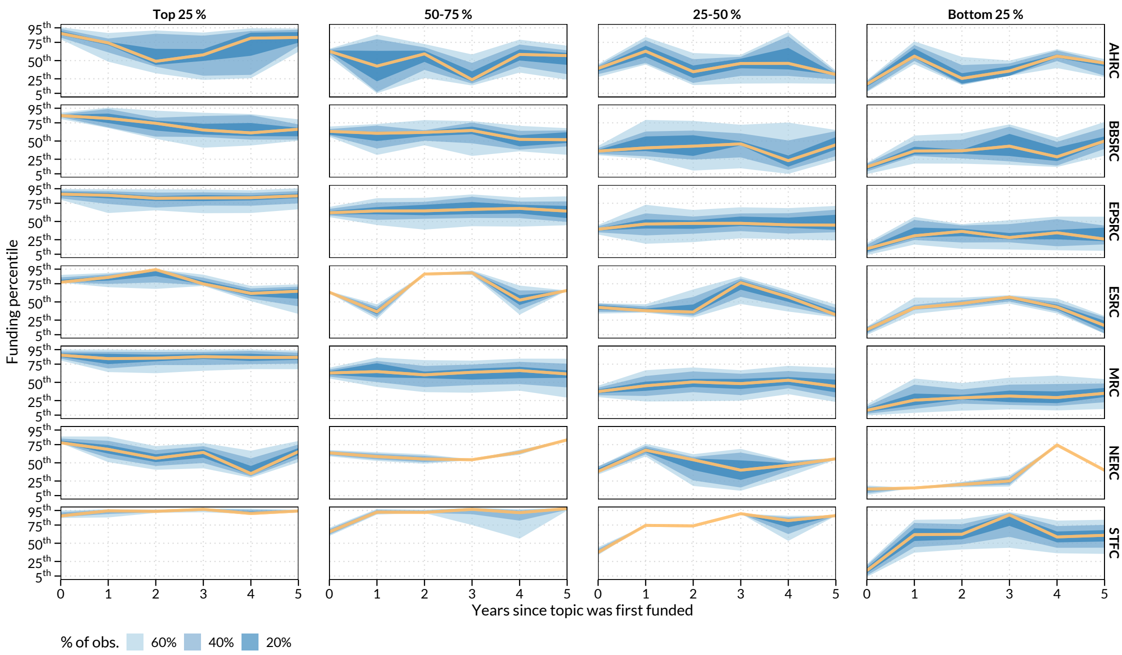


Figure S7.5: Subsequent funding percentiles for topics in each research council. Groups correspond to 25 % increments of the funding distribution when topics were first funded.

Table S7.1: Summary of grant samples

Prior specification	Prior
1	$\beta_k \sim N(0,3), \beta_0 it \sim N(0,20)$
2	$\beta_k \sim N(0,5), \beta_0 it \sim N(0,20)$
3	$\beta_k \sim N(0,7), \beta_0 it \sim N(0,20)$
4	$\beta_k \sim N(0,3), \beta_0 it \sim N(0,40)$
5	$\beta_k \sim N(0,5), \beta_0 it \sim N(0,60)$
6	$\beta_k \sim N(0,7), \beta_0 it \sim N(0,80)$
7	$\beta_k \sim Cauchy(0,5), \beta_0 it \sim Cauchy(0,20)$
8	$\beta_k \sim Cauchy(0,7.5), \beta_0 it \sim Cauchy(0,20)$
9	$\beta_k \sim Cauchy(0,10), \beta_0 it \sim Cauchy(0,20)$
10	$\beta_k \sim Cauchy(0,5), \beta_0 it \sim Cauchy(0,40)$
11	$\beta_k \sim Cauchy(0,7.5), \beta_0 it \sim Cauchy(0,60)$
12	$\beta_k \sim Cauchy(0,10), \beta_0 it \sim Cauchy(0,80)$

7.6.4 Regression models of cumulative advantage

For the regression model of cumulative advantage, I specify a prior for both the effects of earlier funding, and the intercept. Priors help constrain the model to avoid fitting noise in the data. In essence, prior distributions of the parameters help regularise the estimates by introducing additional information about the likely parameter space, which constrains the model procedure through a reduction of variance. However, the lesser variance comes at a cost of higher bias compared to frequentist unbiased estimates. For all priors, I centre the prior distribution around zero to reflect an initial scepticism towards an effect, but vary the type of prior distribution and the scale parameter to allow for more or less regularisation. For all models, I standardised the input variables, the lagged funding levels, by centering them on the average and dividing by one standard deviation. This seemed reasonable, as the lagged funding levels are distributed in a very skewed fashion, and putting the variables on a standardised scale reduces the variance in them and facilitate easier convergence of the model. I show that the posterior inferences derived from the model presented in the main text is robust to a reasonable alternative prior specification. Table S7.1 gives the twelve prior specifications, and Figure S7.6 shows the main coefficients for each model, with 95 % credible intervals. Table S2 provides the posterior means and quantiles for the model reported in the main text.

Table S7.2: Bayesian model of path dependence in funding

	Estimate	Posterior distribution quantiles			
	β_k	2.5th	25th	75th	95th
Predicting level of funding					
β_{0it}	13.72	13.65	13.70	13.74	13.78
One year lag	0.14	0.08	0.12	0.16	0.20
Two years lag	0.16	0.11	0.15	0.18	0.22
Three years lag	0.15	0.10	0.13	0.17	0.20
Four years lag	0.22	0.17	0.2	0.24	0.27
Five years lag	0.11	0.06	0.82	0.86	0.16
Std. dev. of gamma	0.84	0.77	0.82	0.86	0.91
Predicting no funding					
β_{0it}	-1.78	-1.89	-1.82	-1.74	-1.67
One year lag	-3.65	-4.00	-3.77	-3.53	-3.30
Two years lag	-0.72	-0.93	-0.79	-0.65	-0.52
Three years lag	-0.09	-0.19	-0.13	-0.05	0.04
Four years lag	0.08	-0.02	0.04	0.11	0.20
Five years lag	0.65	0.48	0.59	0.71	0.81
Std. dev. of gamma	0.27	0.10	0.23	0.32	0.39

Note: All estimates are on a standardised and mean-centered scale. Estimates of funding levels are on a log-scale, while estimates for predicting zero funding are on the logit scale

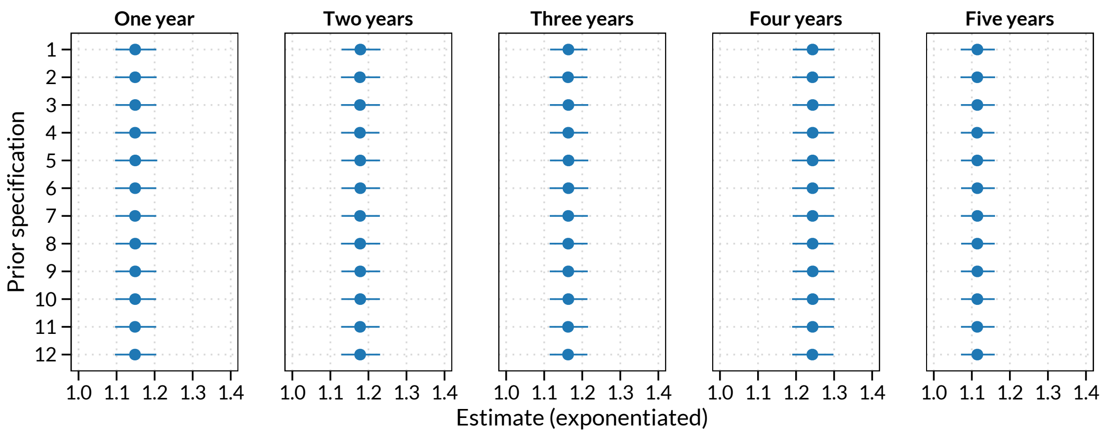


Figure S7.6: Posterior means from twelve different models. Prior specification corresponds to table S7.1

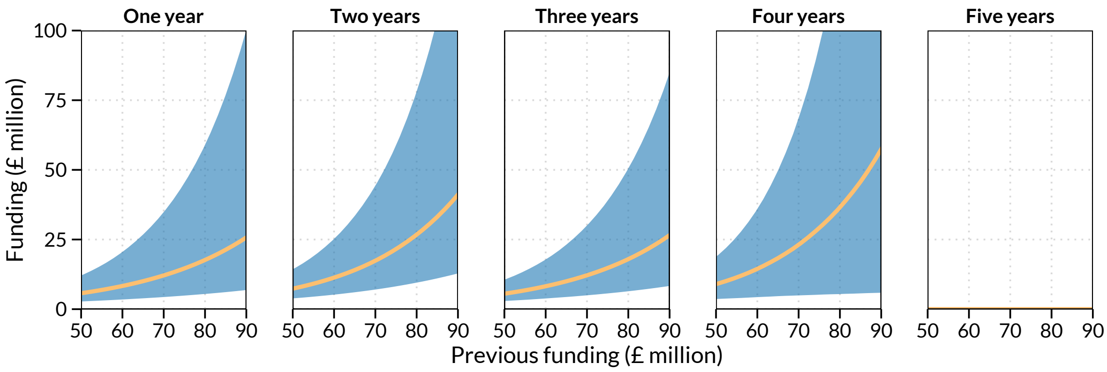


Figure S7.7: Average funding amounts for topics conditional on previous funding amounts. Each panel shows the population average posterior fit for previous funding one to five years before. Posterior fits are based on model 3 in table S7.2 with 95 % credible intervals.

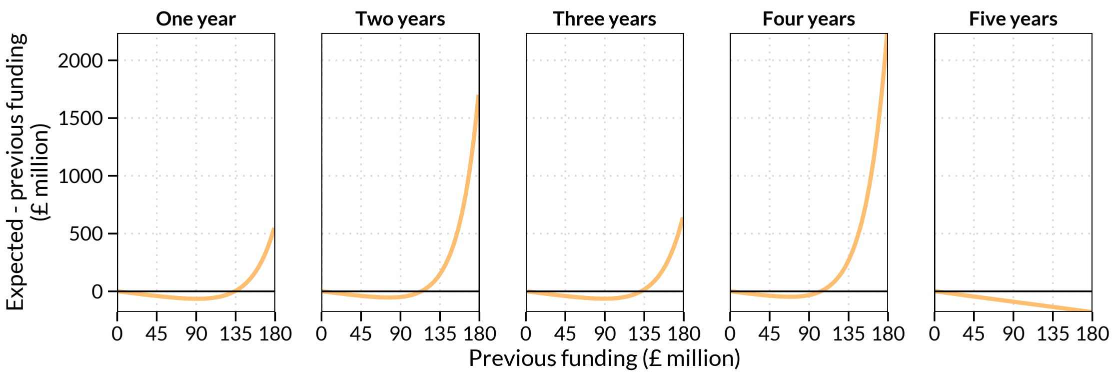


Figure S7.8: Expected mean funding amount minus previous funding amount. Each panel shows the difference between the expected amount of funding in a given year (population average posterior fit) and previous funding one to five years before. Posterior fits are based on model 3 in table S7.2 with 95 % credible intervals.

8. Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers

Emil Bargmann Madsen and Kaare Aagaard[1]

The degree of concentration in research funding has long been a principal matter of contention in science policy. Strong concentration has been seen as a tool for optimizing and focusing research investments but also as a damaging path towards hypercompetition, diminished diversity, and conservative topic selection. While several studies have documented funding concentration linked to individual funding organizations, few have looked at funding concentration from a systemic perspective. In this article, we examine nearly 20,000 competitive grants allocated by 15 major Danish research funders. Our results show a strongly skewed allocation of funding towards a small elite of individual researchers, and towards a select group of research areas and topics. We discuss potential drivers and highlight that funding concentration likely results from a complex interplay between funders' overlapping priorities, excellence-dominated evaluation criteria, and lack of coordination between both public and private research funding bodies

[1] Danish Centre for Studies in Research and Research Policy, Deptartment of Political Science, Aarhus University

RESEARCH ARTICLE

Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers

Emil Bargmann Madsen and Kaare Aagaard

Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University, Bartholins allé 7, 8000 Aarhus C

an open access journal

Citation: Madsen, E. B., & Aagaard, K. (2020). Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers. *Quantitative Science Studies*, *1*(3), 1159–1181. https://doi.org/10.1162/qss_a_00077

DOI: https://doi.org/10.1162/qss_a_00077

Received: 01 April 2020
Accepted: 23 May 2020

Corresponding Author:
Emil Bargmann Madsen
ebm@ps.au.dk

Handling Editor:
Ludo Waltman

Keywords: diversity, funding concentration, research funding, research policy, research priorities

ABSTRACT

The degree of concentration in research funding has long been a principal matter of contention in science policy. Strong concentration has been seen as a tool for optimizing and focusing research investments but also as a damaging path towards hypercompetition, diminished diversity, and conservative topic selection. While several studies have documented funding concentration linked to individual funding organizations, few have looked at funding concentration from a systemic perspective. In this article, we examine nearly 20,000 competitive grants allocated by 15 major Danish research funders. Our results show a strongly skewed allocation of funding towards a small elite of individual researchers, and towards a select group of research areas and topics. We discuss potential drivers and highlight that funding concentration likely results from a complex interplay between funders' overlapping priorities, excellence-dominated evaluation criteria, and lack of coordination between both public and private research funding bodies.

1. INTRODUCTION

Across countries, research funding systems have undergone major changes during recent decades. As funding plays an influential role in the governance of contemporary science, changes in allocation mechanisms are assumed to affect the scope, content, direction, and impact of public research (Gläser & Velarde, 2018; Sörlin, 2007). Such changes are in turn also likely to influence the distribution of funding across individuals and topics and hence the overall degree of funding concentration.

Current evidence suggests that funding concentration has in fact increased as a result of these changes and that there may be some advantages, but also important downsides, linked to high degrees of selectivity in the distribution of funding (Aagaard, Nielsen, & Kladakis, 2020). However, the exact patterns and the potential drivers at aggregated levels are still underexplored.

On the one hand, more pervasive competition, increased performance orientation, stronger emphasis on excellence, and higher reliance on project funding have, across countries, been seen as essential means to optimize the returns on public investments in science (Wang, Lee, & Walsh, 2018). These developments are often highlighted as factors that are likely to increase concentration (Aagaard et al., 2020). On the other hand, funding landscapes have also become more diverse and heterogeneous during this period, with research depending on, but also made possible through, many different extramural sources, such as public research councils, charities, private

The MIT Press

foundations, and firms (Whitley, Gläser, & Laudel, 2018, p. 110). A more diverse set of funders, both public and private, could be expected to ensure a more balanced and diverse distribution of funds, because priorities may differ across funding organizations and funding instruments.

However, how this actually plays out in broader settings, where multiple funders interact, is unclear. Existing studies on concentration have so far largely neglected to assess patterns, drivers, and consequences at the systemic level. Instead, most studies focus on the concentration of funding, to both individuals and research areas, by analyzing grants from only one funder at a time: for example, the National Institutes of Health (Best, 2012; Bowen & Casadevall, 2015; Hegde & Mowery, 2008; Hegde & Sampat, 2015; Katz & Matter, 2019; Li & Agha, 2015; Manton, Gu, et al., 2009; Stoeger, Gerlach, et al., 2018; Wahls, 2018b), the National Science Foundation (Cole, Cole, & Simon, 1981), the UK Engineering and Physical Sciences Research Council (Ma, Mondragón, & Latora, 2015), the Australian Research Council (Bromham, Dinnage, & Hua, 2016), or the Veni, Vidi, Vici Program in the Netherlands (Bol, de Vaan, & van de Rijt, 2018).

While these and other studies provide interesting findings, they are unable to address the combined effects of multiple funders. Hence, although the need for comprehensive knowledge about funding priorities across funders is often stressed, the issue has seldom been examined in depth.

We improve on these studies by analyzing a set of almost 20,000 grants awarded by 15 Danish research funders between 2004 and 2016. The data encompass almost all major research funding bodies in Denmark. To our knowledge, our study is the first empirical study of the degree of concentration of research funding both for individual researchers and for fine-grained research topics/disciplines considering the majority of relevant funders within a country. From this outset we ask the following research question:

What patterns of concentration of research funding can be observed across individuals and topics when the majority of funders within a national system are investigated in concert?

More specifically, we examine the following questions:

- How concentrated is the research funding distributed among individuals?
- Are funding distributions markedly stratified in terms of gender?
- What do the funding configurations of top funded researchers look like?
- How is the funding of the top 100 most funded researchers distributed among universities?
- How concentrated is the distribution of funding across research topics?
- How is the research funding distributed across disease-specific research areas?
- To what extent do the observed patterns match measures of disease burdens (which can be seen as a partial measure of societal demand)?

In addition to providing information on these research questions, the analysis also raises a more tentative discussion of the potential drivers of the observed developments.

The paper proceeds by briefly discussing earlier findings on concentration and diversity in research funding in Section 2. We then outline how we combine detailed information on individual grants and grantees coupled with publication data to answer the questions posed above. In Section 4, we first show the degree of funding concentration in Denmark at the individual level and examine the funding configurations of the top funded researchers. Subsequently we proceed by investigating thematic concentration, first at an aggregated level and second more detailed by focusing on disease-specific research. Here we match the observed patterns with measures of disease burdens to examine whether concentration are driven by societal demands. Based on the total set of findings, we discuss potential drivers of concentration in Section 5. In the final section, we highlight our main findings, discuss the pros and cons of concentration, and derive a series of implications for funding policy.

2. PREVIOUS STUDIES OF CONCENTRATION IN RESEARCH FUNDING

Concentration of research funding has become a key part of the science policy debate. This is due to multiple factors, such as efficiency concerns, austerity policies, and a general trend towards greater reliance on grant funding, which together have sparked discussions on how best to support scientific discovery. Nonetheless, the systematic, empirically based literature on trends in concentration of grant funding is still in its infancy (Aagaard et al., 2020).

Although fragmented, the available scholarly literature in general appears to document a high degree of concentration of grant funding at the individual researcher or group levels. The majority of these studies rely on grant information from the US National Institutes of Health (NIH), and generally show that around 10% of grantees receive around 40% of all allocated funding (Katz & Matter, 2019; Peifer, 2017; Wahls, 2018a, 2018b). While some concentration of grant money consistently has been a part of the system, Katz and Matter (2019) document that concentration has significantly increased over the past 30 years. In 1985, the top 20% funded researchers accumulated around twice the amount awarded to the bottom 20%, but by 2010 this had increased to 3.5 times (Katz & Matter, 2019, pp. 10–11). Similar developments have taken place elsewhere. In the UK's Engineering and Physical Science Research Council (EPSRC), 8% of researchers receive 50% of all funding, and also here the level of concentration rose steeply between 1985 and 2013 (Ma et al., 2015). Along the same lines, two studies focusing on Quebec show how 20% of researchers amass 80% or more of total funding across a broad range of disciplines (Larivière, Macaluso, et al., 2010; Mongeon, Brodeur, et al., 2016).

Another common finding is an apparent lack of diversity extending to both the type of researcher and the type of research funded. At the individual level, the high concentration of funding results in a bias towards researchers from prestigious institutions. In the NIH case, more prestigious institutions receive 240% more funding per grantee compared to less prestigious institutions, have a 65% higher success rate, and receive 50% larger award sizes on average (Wahls, 2018a). The lack of diversity in affiliation is also evident in the EPSRC, where a small group of universities form a "rich club," which attracts the bulk of funding, and act as central brokers in collaborations between different universities (Ma et al., 2015, p. 14763). Similarly, high degrees of concentration also seem to reinforce gender biases in grant competitions. Findings from NIH grant competitions show that women are underrepresented research project grantees, have higher risk of not transitioning to the next level, and often transition much later in their career (Lerchenmueller & Sorenson, 2018, p. 1011).

A second issue related to funding concentration is the possible impact on topic diversity. Observers have long worried that increasing competition for resources, paired with low success rates, incentivize applicants to be more conservative and mainstream oriented in the type of research they apply for (Alberts, Kirschner, et al., 2014). On a system-wide level, this may at the same time lead funders to invest in a narrow set of topics where capacity already exists and where the probability of obtaining results is high. Studies of biomedical research funding again seem to confirm this notion. Funding from the NIH tends to flow towards research into a select group of diseases, often matching allocation from previous years (Yao, Li, et al., 2015). In 1996, 10% of diseases received 43% of the funding budget, with AIDS and breast cancer amassing 36% in total (Gross, Anderson, & Powe, 1999, p. 1882). By 2006, 10% of all disease areas still amassed around 40%, with AIDS and diabetes still being the top funded areas (Gillum, Gouveia, et al., 2011, p. 2). Certain genes are also more likely to receive attention, with up to 75% of projects funded by the NIH relating to only 5% of human protein-coding genes. This affects individual incentives for choosing genes to study, as the probability of becoming a principal investigator (PI) is 10% for researchers studying the least studied genes compared to 25% for those who studying the mainstream (Stoeger et al., 2018, p. 8).

More generally, funding also appears to skew towards already well-funded topics. Outside the biomedical field, a recent study of grant funding in the United States shows a considerably skewed distribution of funding over topics (Klavans & Boyack, 2017). This concentration of funding correlates well with the prominence of topics in the scholarly literature but also with the amount of funding a topic has previously attained. By itself, a topic's scholarly prominence in one time period explains over one third of the funding variation in the subsequent period (Klavans & Boyack, 2017, pp. 1167–1169).

While topic skew signifies a thematic concentration of funding, studies also show a concentration of approaches in the way in which these topics are studied. This is visible in medical research, where different stages of the research pipeline receive different levels of funding. Funding from Danish research councils has been skewed towards translational research (39%) and basic science (21%), while clinical trials only received around 10% of all funding between 2012 and 2016 (Rygård, Kjær, & Perner, 2018). Similarly, focusing on infectious disease research, Head, Fitchett, et al. (2016, p. 185) find that 59.4% of funding in the UK is awarded to preclinical research, while funding for phase I–III trials amounts to 5.6%.

Finally, the societal impact of a disease appears to be moderately related to funding intensity (Gillum et al., 2011; Gross et al., 1999). Recent studies actually find a poor match between estimated health needs and funding distributions (Jones & Wilsdon, 2018, pp. 19–20; UK Clinical Research Collaboration, 2015, p. 49), and this weak link disappears entirely when funding across longer time periods is taken into consideration (Yao et al., 2015, p. 810).

In sum, a growing literature shows that funding tends to be highly concentrated among few researchers, topics, and approaches. However, the available studies are skewed towards a limited number of disciplines and countries and most importantly, there is very limited evidence of such patterns at the systemic level. Furthermore, empirically based studies of drivers of concentration are also still lacking. These issues are examined in Sections 4 and 5.

3. METHODS AND DATA

3.1. Data on Competition-Based Grants

The present study relies on a database of 19,399 research grants awarded by 15 public and private Danish research funders over a period of 13 years from 2004 to 2016. The group of funders encompass both public research councils, private foundations with corporate interests, and non-profit foundations or societies. In addition, ERC grants are also included, as in a Danish context they are perceived as a central part of available funding sources and as some of the most competitive and prestigious grants. The total awarded grants amount to 52.9 billion kroner (€7.08 billion or US$7.7 billion) distributed to 7,539 grantees. It is, however, important to note that this sample is not a complete picture of all grants awarded by Danish research funders within the period. Many funders award money to projects that are not strictly research-related or aimed at producing a traditional scientific output. For that reason, we have excluded all grants aimed at training or education, as well as grants with a monetary value below 50,000 kroner (€6,700 or US $7,400). The bulk of these grants are small travel grants or grants to students ("skolarstipendier").

For all grants, we standardized the grant sums to be in kroner and current prices. For grants with a running amount paid each year, we summed these to one grant sum centered on the grant year. In addition, it is also important to highlight that grants are linked to the main recipient only (typically the PI). For this and other reasons the database has less comprehensive coverage of strategic and innovation-related funding, as these types of grants are often awarded to a consortium of firms and research organizations without a single discernible PI. Consequently, some underrepresentation of these sources is likely. Furthermore, the focus on PIs only also means that our analyses cannot account for how the funding is spread across researchers after being awarded to a PI.

After establishing the database, grant recipients were manually disambiguated and linked to grants. For common combinations of names, the grantee was identified through a combination of data on their institutional affiliation and primary research area. If not successfully disambiguated through this, grantees were considered separate individuals even if names matched. Throughout the disambiguation process, we also established the gender of the grantee for 97% of all individuals. In addition, a number of infrastructure grants and PhD and postdoctoral block grants were awarded to deans or rectors. These grants cannot reasonably be treated as grants to individuals. They were instead assigned to an institution and kept out of the individual level analyses.

Despite these limitations, the database covers a large part of all external funding awarded to Danish researchers from 2004 to 2016. Comparing the data from 2012–2014 to self-reported sums spent by the funders in that period, coverage of around 2/3 of all grant sums is found (UFM, 2016). These are likely not completely overlapping, but shows that the coverage of central funding organs is consistent across time. Although not all funders had equally accurate data of their grant history, we believe that the overall trends of the aggregated distribution of research funding are quite valid. Hence, while the data may not accurately depict the absolute amount of funding awarded, they provide an important insight about the relative distribution of scarce resources in the Danish system. Again, however, there is a caveat linked to strategic and innovation-oriented research funding, where both the absolute and relative distributions are more uncertain.

3.2. Connecting Grants and Outputs

Few Danish funding organizations have consistent ways of classifying the research topics funded by their grants. Hence, to ascertain how funding is distributed across different types of research we rely on journal articles and reviews published by the grantees. To link each grantee to their publications, we use a combination of automated and manual matching, by comparing names, email addresses and institutional affiliations in the CWTS in-house version of the Web of Science (WoS). Using this method, we matched 5,773 grantees and 14,103 grants, to their publication profiles. Together, these grants account for around 80% of the combined funding amounts attributed to individual principal investigators. Most of the nonmatched grants were awarded to research within the humanities and social sciences, which have less extensive coverage in the WoS (Mongeon & Paul-Hus, 2016). The full and WoS-matched samples are predominantly similar in their composition. The WoS-matched sample has 2% fewer female grantees, and the grants tend to be a little larger and have a higher percentage of research grants (compared with, for example, postdoctoral grants). The differences are, however, negligible. Tables A1 and A2 in the Appendix provide a more detailed comparison of the two samples.

In the analyses below, we use the publications of each grantee to attribute them to a research area or a specific research topic. This entails some methodological choices. We assume that a principal investigator's publications reveal the underlying topic, or mix of topics, of the grant. We also assume that the topic mix has not changed much from grant proposal to publication output. However, using the publication output may also carry the advantage that we are assessing the type of research that a grant actually resulted in, instead of just the intended results. Furthermore, because publications may span multiple areas and topics, we may better represent cross-disciplinary research grants. A grantee's publication output will not be a perfect representation of the topicality in a grant, but is expected to yield a reasonably accurate estimate of topic spread at the aggregate level. We only consider the publications 0–4 years after a grant was received. This approach leaves us with 63.3% (12,269) of all the collected grants.

To study the concentration of funding in different research areas and research topics we rely on two categorization schemes. To categorize each grant within a research area, we use the WoS OECD Category Scheme of 39 research areas categories.[1] The categories are based on the 250 journal classifications in the WoS database. For publications in multidisciplinary journals, we investigated the most common category of all references in a paper and assigned it to this.

In the article, we furthermore conduct a small case study of disease-specific research. The Danish system, in addition to public funders, also includes private research funding foundations tied to pharmaceutical companies. Disease-related funding then provides the opportunity to compare how the priorities of public and private funders overlap. Through publications, we were able to match 4,108 grants, spanning 15 funders, to one or more of the 134 disease areas used by the WHO in assessing global burden of diseases.[2] The classification of papers into disease areas is based on data from Yegros-Yegros, Klippe, et al. (2019), which translates medical subject headings from the PubMed database into ICD-10 (International Classification of Diseases) categories, and finally into disease areas.

For each grant, we weigh both the grant sum and the grant itself according to the proportion of publications within each research area or disease category. For example, a grant of 1 million kroner with three publications within chemistry (60%) and two publications within nanotechnology (40%) is divided so that 0.6 grants and 600,000 kroner are attributed to chemistry, and 0.4 grants and 400,000 kroner are attributed to nanotechnology. Finally, following previous studies, we use data on the burden of disease categories to investigate how funding aligns with (imperfect) measures of societal health needs (Cassi, Lahatte, et al., 2017; Evans, Shim, & Ioannidis, 2014; Gillum et al., 2011; Gross et al., 1999; Yao et al., 2015; Yegros-Yegros et al., 2019). This is done by comparing research output for a disease to its burden in Disability-Adjusted Life Years (DALYs). DALYs are a common measure of disease burden that estimates the equivalent number of healthy years lost due to disability or early death (Gillum et al., 2011). It sums both the Years of Life Lost (YLL) and Years Lived with Disability (YLD) estimates for each disease. We match DALY estimates from the 2017 Global Burden of Disease Study (GBD 2017 DALYs and HALE Collaborators, 2018) to the 134 disease categories from the WHO, and convert these to relative burdens by summing DALYs from all diseases and calculating the share of total burden attributed to each disease.

In this study, we use DALYs as a rough proxy for the societal relevance or importance of a disease category. However, while widely used, DALYs are also a heavily criticized health metric. A comprehensive critique of the DALY measure is beyond our scope, but we highlight a few salient issues. From a methodological standpoint, the DALY calculation relies on a set of strong assumptions of which not all necessarily are agreed upon (Anand & Hanson, 1997; Parks, 2014). The calculations are furthermore characterized by a great deal of uncertainty and can be an inaccurate measure of disability in high mortality regions, for example, because the underlying data are rough estimates or simply missing (Parks, 2014). DALYs also have several shortcomings from a conceptual perspective. First, by pooling YLL and YLD estimates,

[1] http://help.prod-incites.com/inCites2Live/filterValuesGroup/researchAreaSchema/oecdCategoryScheme.html

[2] https://www.who.int/gho/mortality_burden_disease/en/

the DALY indicator conflates death and disability into one scale, with premature death as an extreme case of disability. Each single case of a disease is also discounted based on the person's age, and case contributions to the DALY estimate thus vary. In sum, the DALY measure relies heavily on a "cost effectiveness" rationale in both how it is calculated and what this implies conceptually. Such a rationale can been seen as an inherent political choice, where other rationales, such as "equity," could be equally or more appropriate. Nevertheless, DALYs can provide a reasonable measure of health burden in Denmark, as we are not comparing health burdens to other contexts and expect underlying data to be reasonably well curated. Our aim is not to argue for a distribution of research funding according to DALYs but rather to contrast the distribution of funding to an indication of societal needs. However, we acknowledge that by comparing burden and funding, there is a risk that we appear to legitimize some of the criticized implications of the DALY.

4. ANALYSIS: CONCENTRATION OF DANISH RESEARCH FUNDING

4.1. The Danish System for Funding Research

The Danish research funding system is pluralistic and consists of many different funding channels, many individual funding programs and a variety of specific funding mechanisms. This organization reflects a need to serve different societal purposes, the involvement of a broad array of sectoral interests, and an aim to underpin a variety of different outcomes. However, in line with the general international development, the Danish system has also become more competitive over time. Most importantly, a development was started in 2006 towards turning the existing 65/35 balance between institutional funding and external funding into an approximately equal 50/50 balance (Danish Government, 2006). Second, a substantial part of this increasing share of competitive funding was channeled into several newly established research funding organizations, which were institutionalized alongside the traditional research council system and the Danish National Research Foundation. Recently, these research-funding organizations have been merged into one, namely the Innovation Fund Denmark (established in 2014). Accordingly, Denmark now has a rather sharply divided research funding system with a clear distinction between response mode funders and more mission- or innovation-oriented funders. In addition, the Danish research system has for many years benefited from a strong and varied sector of nonprofit and private research funding foundations. This has been the case for biomedical research in particular, but also most other areas have at least to some extent benefited from these nonpublic funding organizations. The importance of the private research foundations has surged over the past 10 years. Funding from private foundations now comprises almost 20% of all research funding financing public research activities, and close to 50% of all competitive funding (Danish Council for Research and Innovation Policy, 2020). The most recent part of this development is, however, not captured in the present analysis, which only covers funding information up until 2016.

4.2. Concentration of Funding at the Individual Level

In the investigation of the concentration of competitively awarded research funding, we first focus on the number of grants and the amount of funding for each individual grantee in the database. Here we observe a significant skew of resources. Among all 7,539 grant recipients, only 3,000 have been awarded two grants, 1,600 more than two, and 600 individual researchers five grants or more during the period. At the absolute top end, 140 PIs have amassed more than 10 grants, and a small selection of these have close to 30. In total, 20% of the grantees were awarded 50% of all grants.

This level of concentration is even more pronounced when focus is shifted to the total amount of funding for each PI. Figure 1 shows the cumulative share of funding for grant recipients as a Lorenz-curve. A straight line would signify total equality in funding amounts, but the graph instead shows significant stratification among researchers. The blue curve shows the distribution for all 7,539 grantees, where the 20% most successful receive around 75% of all funding, while the remaining 80% share the remaining 25%. This marked concentration of funding at the individual level does not take into account the number of researchers who have either not applied for or not won external research funding. To estimate the concentration of funding across the entire population of Danish researchers, we therefore identified 21,000 publishing scientists with a Danish affiliation and at least five publications in the 12-year period in WoS. As a back-of-the-envelope calculation, this yields a tail of more than 13,000 researchers with no external funding. In Figure 1, we have included three more or less conservative estimates of the number of researchers: 15,000, 20,000, and 25,000. Even when using the most conservative benchmark of 15,000 active researchers over this period, we see a very high degree of concentration. The top 10% attracts around 75% of all funding, while the top 20% amass just short of 90% of the total amount.

The stark degree of concentration can also be shown in terms of the average amount of funding awarded to each grantee. Figure 2 shows the mean funding amount for six ranked groups of researchers. The top 100 researchers each have an average of more than 90 million kroner per person. The group from 101–500 have an average of around 30 million per person, while the 501–2,000 group on average have a little less than 10 million kroner per person each at their disposal. Further down the distribution the average funding amount per person is negligible or nonexistent. However, as highlighted in Section 3, many of these researchers may indirectly benefit from grants as team members of successful PIs.

Besides the average funding amounts, Figure 2 also shows a clear gender bias. Only 16 women are among the top 100 PIs, and the next group of 400 researchers included only 60 women. Hence, as the average amount of funding drops, the proportion of female PIs instead rises.

This pronounced lack of gender diversity does not reflect the population of Danish researchers. Figure 3 shows the proportion of female and male researchers in the population[3] as full-time equivalents (FTE), and compares these to the proportions of female grantees, the proportion of grants awarded to female grantees, and the proportion of funding amounts awarded to female grantees. Overall, 40% of Danish researchers are female, while this is the case for only 34% of the grantees. However, when considering the distribution of grants and funding amounts, only 29% of all grants have a female PI, and only 22% of all funding is allocated to a female PI. Dividing these distributions into broad research areas reveals that lack of diversity in the system trails the lack of diversity in the Natural, Technical, and Health Sciences. While the proportion of female researchers is fairly similar in the Humanities and Social Sciences, the increased narrowing from FTE to funding amounts is pronounced in the harder sciences. This also coincides with the distribution of female researchers at different career stages. In the Humanities and Social Sciences, the proportion of female associate professors is twice as large as in the hard sciences, and the same is true for full professors.

It can also be observed that the grantees' career stage seems to be an additional factor leading to stratification. As an imperfect measure of each grantee's "academic age," we calculated the maximum number of years from their latest grant to their earliest publication for 4,047

[3] Average for 2007–2015 based on numbers from Statistics Denmark.

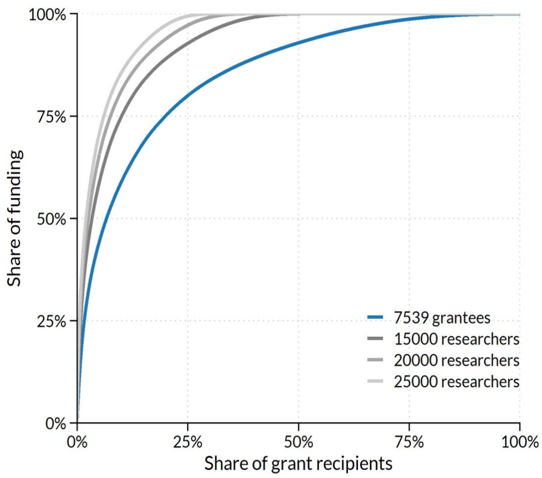


Figure 1. Cumulative share of funding for grantees and estimated population of researchers.

researchers with 11,047 grants. A researcher with an academic age of 5 years has accumulated a median of 1.5 million kroner in grant funding, while a researcher with an academic age of 10 years generally accumulates 4.2 million. Moreover, the distribution of funding becomes much wider for higher academic ages. Table A3 and Figure A1 in the Appendix show that 50% of cumulative funding is contained within the interval [1,995,208–10,331,734] for 10 years of academic age. For 15 years, the interval is [3,959,349–20,939,255]. These numbers should, however, be interpreted with some caution. A significant number of grantees were not matched to previous publications, and while these tend to have lower levels of funding, the distribution within this group varies a lot (see Figure A1).

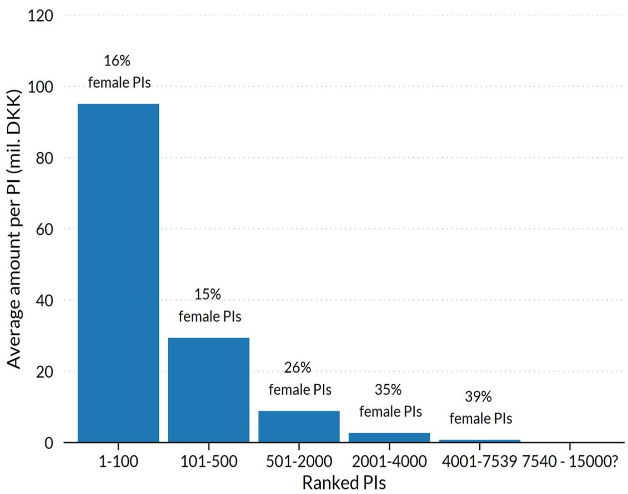


Figure 2. Average amount of funding for grantees and estimated population of researchers.

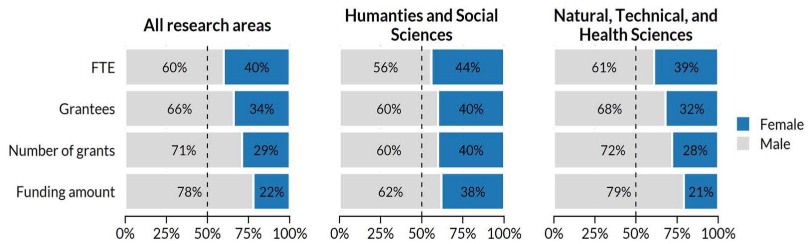


Figure 3. Proportion of female researchers among population (FTE) and grantees, and proportion of grants and funding amounts for female researchers.

4.3. Funding Configurations of Top Funded Researchers

Returning to the small group of the top 100 funded researchers, we also note that this group of grantees separate themselves from the rest by the configurations of their funding. First, the top 100 tend to win grants with higher monetary values. Figure 4 shows the distribution of all grants awarded to the top 100 and the remaining group of grantees, with the median funding amount in black. The top 100 funded researchers tend to have a higher median grant amount. Where the remaining grantees tend to win grants of around one million DKK, the central tendency for the top 100 is 2.7 times that. However, the distribution also shows that the top 100 combines both very large grants with numerous smaller ones, as is evident from the large overlap between the distributions. On average, the top 100 receive almost seven grants and a median of 5.5, while the remaining group receive an average of two grants but a median of one. The top 100 also attract funding from more funders, with an average of 3.2 compared to 1.3.

Second, Figure 5 shows the configuration of unique public and private funding sources among the top 100 funded PIs and the remainder. Within the top 100, the diversity of funding sources is much greater than in the remaining group, with an average of 1.8 private funding sources and 2.2 public, compared to 1.2 and 1.1 among the rest of the grantees. However, the most interesting pattern emerges from the configuration of funding from different sources. Outside of the top 100, 88% of grantees are funded by no more than one private and one public funder. In fact, 77% are only funded by one funder of either domain. In the group of top funded researchers, the configuration distribution is much more spread out: 47% of these grantees are funded by two or more funders from both domains. In essence, the configuration of funding sources is much more diverse among the well-funded scientists, and much more concentrated outside the top 100 group. The biggest concentration within the top 100 is, however, still centered on one public funder (18% of PIs). This reflects a group of grantees whose main source of funding is a large

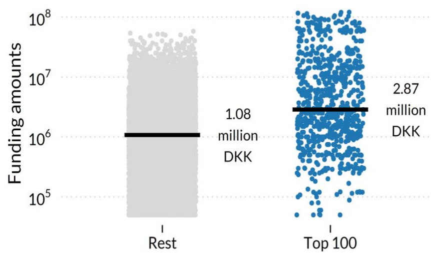


Figure 4. Distribution of funding amounts (median funding amount in black).

Figure 5. Configuration of distinct funding sources.

Excellence Centre grant from the Danish National Research Foundation. These grants exceed 20 million kroner, but are often much larger, with a median of 65.75 million kroner.

Finally, with regard to the top 100 it is also interesting to examine to what extent the grant income of these researchers is concentrated within a few universities. As highlighted in Section 2, studies from the United States and United Kingdom have found a strong bias towards a small group of elite universities. As can be seen from Figure 6, however, this is not the case in Denmark. While the largest and most research-intensive universities obviously account for the biggest share of the total amount of grant income allocated to the top 100, they do not receive a disproportionate share. This likely reflects a Danish university system that is much less stratified than Anglo-Saxon systems and the fact that the relatively low number of universities all are research universities.

4.4. Concentration of Funding on Research Topics

A subsequent question is how these allocation patterns influence the systemic diversity of research areas and topics. Figure 7 shows a disciplinary map of 12,269 of the grants in our data set. These are the grants we could link to one or more publication(s) in the 4 years succeeding a grant. Each bubble or node represents a research area, based on 39 categories from the OECD Frascati manual, and the links between them are the citation traffic between each area. Overall, the figure shows a marked skewness of competitive funding towards the biological and medical sciences.

The biological research areas have, in the examined period, attracted 6.9 billion kroner, followed by 4.6 billion kroner for basic medical research and 2.8 billion kroner to clinical

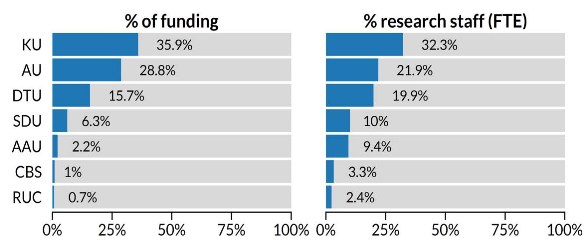


Figure 6. Institutional distribution of funding in top 100 compared to percentage of total research staff in university sector.

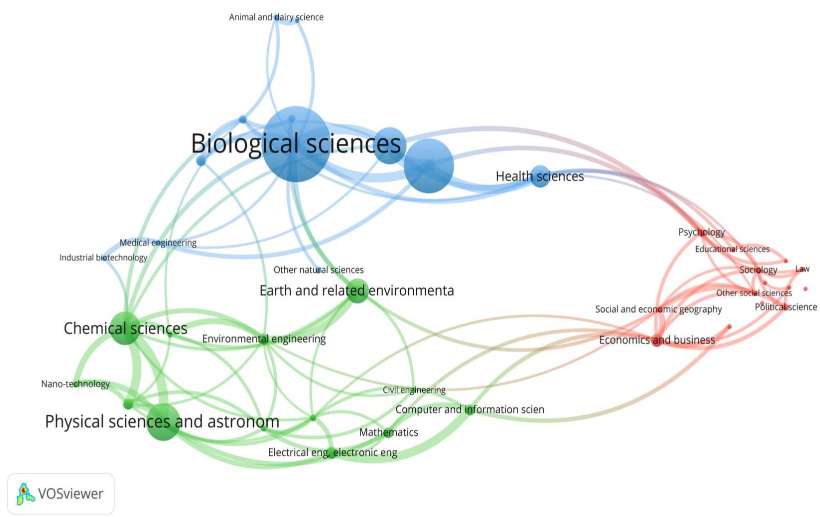


Figure 7. Funded research areas and their citation relations.

research. This picture, however, also illustrates some of the difficulties of categorizing the type of research funded. The biological sciences, as they are defined by the OECD, comprise a broad mix of subdisciplines spanning human, animal, and plant biology. In our disciplinary map, the biological sciences are closely related to the medical research fields, as seen from the strong citation link between these. The OECD categories may then obscure that a substantial amount of external funding is directed towards areas with a close affinity to basic medical research, such as microbiology, biochemistry, molecular biology, genetics, and virology. It is noteworthy that this degree of concentration is relatively similar across different types of funders.

Figure 8 shows the distributions of funding from public and private funders across the top 20 funded OECD areas. It is interesting to note that both types of funders largely prioritize

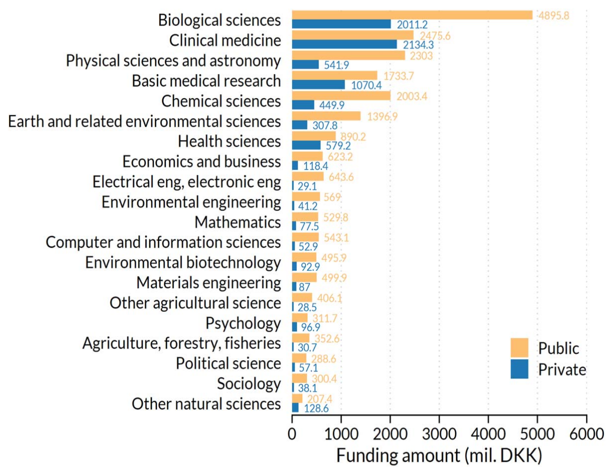


Figure 8. Public and private funding of research areas.

similar, and relatively few, areas. Again, the biological sciences and clinical medicine are the top funded disciplines: The top seven areas are all within natural or biomedical sciences and amass around 70% of total funding. We also see that 25% of areas get 73% of all funding from public funders, and 86% from private funders. So even though both types of funders prioritize a small set of areas, some differences still exist between their respective funding portfolios. While public funders direct the majority of their grants towards biology, physics, chemistry, and clinical and basic medicine, private funders focus primarily on biology and clinical medicine. This is not surprising given the composition of private Danish foundations, where the largest ones are linked to the pharmaceutical industry.

4.5. Concentration of Research Funding on Disease Specific Topics

At the aggregate level, the distribution of both public and private funding flows towards similar types of research, as shown above. However, it is not clear whether this is a case of private priorities that spill over into public funding, or simply a natural focus on the most important and promising areas from both types of funders. To investigate this further, we delve deeper into the distribution of funding within the biomedical area and compare funding of 134 disease categories with measures of their societal health burden.

Figure 9 shows the average relative burden, in the form of a disease's share of total DALYs against the relative funding amount for that disease. For the measure of relative disease burden, we have averaged the number of DALYs for a disease across 2004–2016, and calculated its share of total DALYs. The first plot is the average relative burden in Denmark, while the second shows the average relative burden at a global scale. In each plot, the diagonal shows a proportional relationship between burden and funding. The majority of diseases attract very little funding and pose a relatively small burden to societal health. A number of diseases are placed below the 45° line, indicating that their relative burden in Denmark is not matched by research investments. One example is back and neck pain, which constitutes 9% of total health burden but under 1% of total funding. Similar patterns apply for chronic obstructive pulmonary disease, ischemic heart disease, and stroke. Another set of diseases above the diagonal receive a higher relative share of funding than their burden share. Diabetes is one

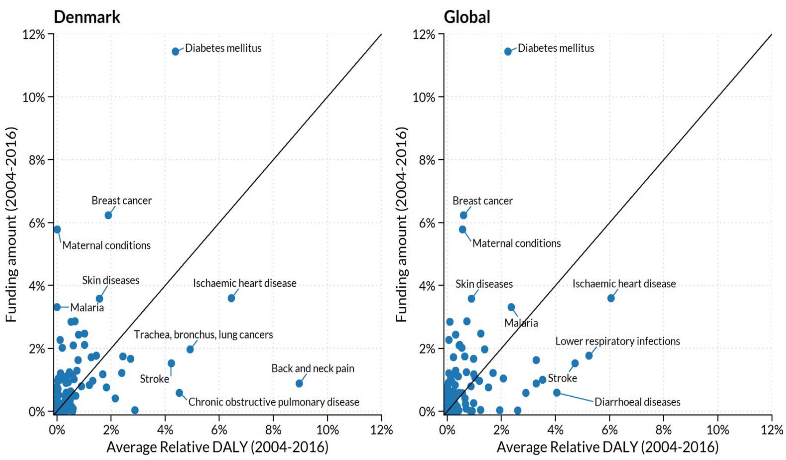


Figure 9. Relative disease burden and funding.

such disease, which receives around 11.8% of all disease-specific funding. Despite having a health burden almost equal to chronic pulmonary disease, diabetes receives 10 times the amount of research funding. On the global scale, a group of diseases moves further towards the diagonal because they pose a larger health burden across the world. This is the case for malaria, diarrheal disease, and HIV/AIDS. In contrast, diseases such as breast cancer and skin diseases appear more "overfinanced" because they primarily affect people in high-income countries (Yegros-Yegros et al., 2019).

The question is why some diseases, such as diabetes, draw so much funding relative to other diseases. Figure 10 may provide some answers to this. We calculated the funding percentile of each of the 134 diseases within the group of private and public funders and plotted these against each other.

The black regression line clearly shows the high correlation between the patterns of allocation by private and public funders with a correlation coefficient of almost 0.9 ($r = 0.892$). Diabetes and breast cancer are the most funded diseases from both private and public funders. Diabetes draws 727 million kroner from public funders and 355.5 million kroner from private funders, while breast cancer was awarded 331.6 million kroner from public funders and 257.7 million kroner from private funders. As both private and public funders tend to focus on a select subset of diseases with high burdens (represented by the size of each bubble), the concentration on a few diseases increases. The figure also shows a great overlap in diseases not prioritized by both types of funders, which indicates both a positive feedback mechanism for high-funded diseases and a negative feedback mechanism for low-funded diseases.

There are, however, also differences in the revealed funding patterns between the two types of funders. A range of neurological and psychiatric diseases are for instance better funded by private funders, and while, for example, Alzheimer's disease and dementia are within the top 10 most burdensome diseases in Denmark, 57.5% of funding for these diseases is from private funders. Likewise, schizophrenia, stroke, ischemic heart disease, and some types of cancer (breast, skin, brain) rely on over 40% of funding from private sources. In contrast, a range of diseases prevalent

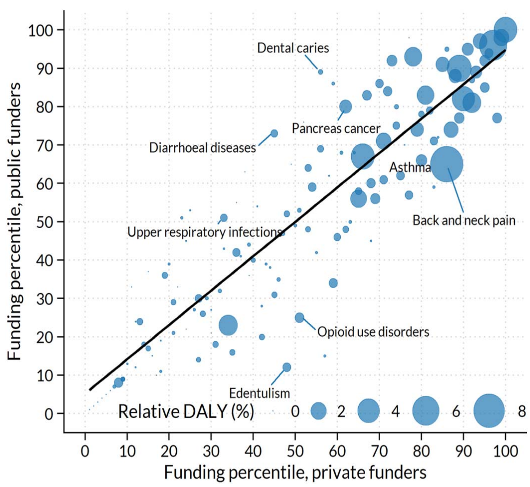


Figure 10. Funding percentiles for 134 disease areas in private and public funding bodies. 0 = lowest, 100 = highest.

in other parts of the world, such as HIV/AIDS, lower respiratory infections, and maternal complications receive only 25–30% of funding from private funding bodies. So even though priorities often line up across funders, some functional division between them appears to exist where private money flows to disease areas relying on new medicine, while public resources are channeled towards well-studied diseases with a need for large-scale prevention interventions.

5. DISCUSSION: DRIVERS OF FUNDING CONCENTRATION

The Danish research funding system produces pronounced concentration with consequences for the research population as a whole, gender equality, and topic selection. This degree of concentration is not least noteworthy, as the Danish system is considered relatively egalitarian.

However, the scholarly literature does not offer convincing explanations of such allocation patterns. A main reason is most likely that no simple explanation exists. Rather, there may be a number of mutually reinforcing causes interacting in various ways. In the following, we highlight potential drivers and relate them to the empirical findings in Section 4. While this approach is insufficient to draw clear causal conclusions, the findings raise important discussions and indicate how different drivers interact in a specific national context. This may in turn inform policy decisions and highlight needs for further comparative research.

5.1. Internal and External Drivers

Concentration is not a new phenomenon. The institutional structure of science has always been biased toward concentration. Time and again, the pronounced inequality and the elitist nature of scientific activity has been demonstrated, most clearly manifested in the highly skewed distribution of productivity and recognition among scientists (Allison, 1980; Allison & Stewart, 1974; Cole, 1970; Fox, 1983; Lotka, 1926; Merton, 1968; Price, 1963; Reskin, 1977; Zuckerman, 1970). Numerous historical examples have showed that even in cases where external pressures were marginal or absent, highly stratified systems have been the rule rather than the exception (Merton, 1968). Hence, some degree of selectivity in the allocation of funding has traditionally been seen as both natural and justified.

However, recent conscious and deliberate policy changes related to the funding and assessment of science may have amplified this inherent bias. Larger grants, increased support for critical mass and emphasis on initiatives to create "world-leading" environments are only some of such measures. Seen from this perspective increased funding concentration is both an intended and desired outcome.

But increased funding concentration may at the same time also be the result of less obvious and less deliberate factors. Two of these more hidden factors seem particularly important. The first concerns the dominant research quality criteria put to use when project proposals are assessed in multiple sites. If different funding agencies operate with relatively uniform criteria based on narrow notions of excellence (typically judged by elite peers and supported by metrics) priorities are likely to be mirrored even across different funders. Hence, if a majority of funders all aim to pick and fund the "best" researchers based on similar excellence-oriented quality criteria, the result is likely to be increased concentration.

This tendency can be further amplified by a second factor: a lack of oversight of allocation decisions made elsewhere in the system. In situations with limited coordination and transparency within and across grant bodies, the result may be higher concentration than any single funder or policy-maker aims for. Even if each single grant decision in isolation is sound, the systemic effects may be undesirable when the majority of funders select using identical parameters, with many funders inadvertently ending up funding the same researchers and the same narrow topics. If this is the case, systems with a broad variety of funders may in fact risk ending up with higher degrees of concentration than systems with more centrally steered allocation decisions.

5.2. Drivers in a Danish Context

The question is to what extent these potential explanations match the findings from Section 4. As we will show in the following, all the above-mentioned factors appear to have been present and active in the Danish case, interacting to amplify concentration. First, the Danish system has experienced deliberate policy attempts to increase the level of competition as well as the degree of selectivity in the allocation of research funding over the past 15 years. As outlined in Section 4, the share of project funding has increased from less than a third of the total research funding to nearly half. Alongside this, grant sizes have grown and success rates have dropped (Aagaard, 2017). An explicit policy aim has here been to try to raise the quality of Danish research through increased competition and by deliberately amplifying central elements in the institutional bias towards concentration.

Over time, the explicit aim of funding only the most excellent ideas has also become much more pronounced across all funders—public as well as private. In this situation, researchers who have high publication and citation performance and already have access to funding have been disproportionally likely to be further rewarded across the board. This is, for instance, reflected in the patterns of most of the top funded Danish researchers' success in winning funding from many different funders. In addition, there are also indications that such self-reinforcing positive feedback loops initially may be set in motion by private foundations. These foundations, with their additional funding and specific priorities, are likely to give some researchers and some topics the upper hand, which subsequently may become amplified by public funding organizations. Hence, when private foundations have specific topical interests and establish and support visible, impactful research groups, this has self-reinforcing positive and negative feedback effects for the system as a whole. This trend may be even further amplified when the most successful grant recipients are subsequently rewarded with additional institutional funding via performance-based internal funding allocation criteria (Aagaard, 2017). In this way, the cycle may continue and perpetuate even further concentration.

Indications that self-reinforcing effects are active in the Danish system can be found not only at the individual level but also at the topical level. As shown in the case study of disease-specific research, there is large overlap in the priorities set by different funders. Even though both public research councils and private foundations focus the majority of their funds on disease areas with a relatively high disease burden, their common focus exacerbates the skew of the resource distribution. Part of the explanation of this overlap in priorities is likely to be an overlap in selection criteria between different funders. But it is also observed that public funding is least concentrated in research closest to patients, and that private sponsors therefore become primary funders of clinical research. These foundations may naturally focus on a narrow set of commercially viable areas (Rygård et al., 2018) and researchers working within these areas may have easier access to funding. Funding may then concentrate even further on both individuals and research topics.

When a preliminary version of this analysis was presented in Denmark in 2019, funders, policy-makers, and stakeholders expressed surprise and concern over the actual degree of concentration on individuals and topics. Similarly, the distribution of funding across research areas also came as a surprise to most actors. Among all involved parties there was general agreement that better oversight and coordination are needed in the Danish system to ensure a more balanced overall investment portfolio. The impression given was also that the degree of selectivity and the actual priorities did not only reflect deliberate decisions made with open eyes. To a large extent the results were also perceived as the result of information asymmetry to the benefit of the most successful researchers. The observed patterns appear in other words to be the result of both intended and unintended mechanisms reinforcing each other. Hence, at least in the case of Denmark, a pluralistic and decentralized funding system does not seem to lead to increased diversity—rather the opposite.

There are, however, two additional factors absent in this analysis, which also need to be taken into account. The first of these is internal to science and relates to the perceived prestige of different topics among researchers, journals, reviewers, and funders. While it is hard to conceptualize and measure prestige at this level of aggregation, it can be assumed that prestige considerations to some extent become internalized in assessment criteria and notions of excellence. Hence, the prestige factor may further contribute to amplifying the trend towards concentration. The second factor is mainly external and relates to grand societal challenges and other broad societal priorities. These priorities will in most cases be made from the top down and may create other dynamics across funders. Still, it is not evident whether societal priorities will amplify or weaken trends towards concentration. The societal priorities may in some cases be in opposition to the more science-internal dynamics. Here, the incentive structure of science, with its emphasis on priority and prestige, may in fact disincentivize scientists to study the most pressing societal problems. While most scientists are likely to be motivated both by making scholarly contributions and solving pressing problems, the dominant disciplinary quality criteria may limit the types of problems they can address. Topic choice is, however, a complicated phenomenon and may be influenced by a host of factors, including expected academic returns, feasibility, possibilities to publish, and costs of doing research (Leisyte & Dee, 2012). So while the possibility to attract funding to some extent may influence topic choices (Gläser & Laudel, 2016; Whitley et al., 2018), researchers in general appear to be hard to influence in terms of research directions (Myers, n.d.). Hence, the most effective instrument may not be specific programs, but rather funders' decisions *not* to fund certain topics (Gläser, 2019).

Societal priorities may thus in some cases lead to alternative research lines that would not otherwise have been followed, resulting in increased diversity, but may also end up funding already existing research lines under new headings. In other cases—as shown in relation to the overlap in de facto priorities between public and private Danish funders—societal or specific sectorial priorities may spark a self-reinforcing positive feedback circle, where successful applicants to mission-oriented calls in turn also improve their chances of receiving more bottom-up oriented funding. But because the present analysis has an underrepresentation of such societally oriented grants it is beyond this study to shed light on these types of dynamics in more detail.

6. CONCLUSION AND PERSPECTIVES

Overall, the empirical examinations presented in this study document a strong concentration of Danish research funding allocated by 15 of the largest Danish funders. These patterns of distribution are found in particular at the individual level, where a small proportion of the research population accounts for a very large share of the funding. In turn, it also results in a skewed gender balance as well as skewness in the relative weight of different research areas and more specialized research topics. This Danish situation appears at least partly to be a result of increased competition for external funding, decreasing success rates, many competing funding organizations using rather uniform excellence criteria, and a general lack of coordination and oversight.

However, the dynamics outlined above are not unique to Denmark but appear widespread and rising in many national funding systems around the world. Hence, similar or even greater funding concentration may be found in other national contexts. In this concluding section, we therefore discuss the pros and cons of further developments towards concentration. The discussion first and foremost draws on a recent systematic literature review carried out by one of the authors of this article and two other colleagues (Aagaard et al., 2020).

Some arguments in the scholarly literature clearly favor at least some degree of concentration. A classical meritocratic argument is that the scientists with the greatest potential to produce (potentially) path-breaking research should be rewarded according to their abilities. Economies of scale, critical mass, and access to expensive instrumentation are other arguments. Funding concentration is furthermore argued to give increased flexibility to researchers, allowing them to take risks and pursue their research process with long time horizons. Finally, positive spillovers to nearby research environments, recruitment, and collaboration effects are also mentioned. These all seem rather strong arguments and yet there are indications—which we return to below—that many of these apparent benefits might be achieved with moderate degrees of concentration without the potential systemically counterproductive effects of overly high concentration.

Arguments in favor of dispersal and diversity highlight in particular that the support of many lines of inquiry spreads risk and increases chances of breakthroughs by allowing for a broader variety of perspectives, interpretations, and predictions. Likewise, chances of serendipity can also be assumed to increase with a multitude of competing approaches. Dispersal is simultaneously perceived to foster resilience in constantly changing research systems, where concentration, on the other hand, can lead to stagnation and reduced systemic adaptability. One reason is that large self-perpetuating research units may reduce the systemic capacity to respond flexibly. Large units may at the same time also turn talented group leaders into "science managers" with little time for research and mentoring and with overly strong incentives and pressure to apply for and obtain ever more resources than can be productively spent. Dispersal on the other hand is seen as supporting a broader knowledge pool, creating absorptive capacity across systems as a whole and underpinning research-based teaching across all disciplines. In doing so, it may also secure a strong future growth layer of early and midcareer researchers and keep a broader group of researchers and students active in research. Finally, dispersal is argued as preferable, as it reduces trends towards hypercompetition and may mitigate a peer review system that is perceived as unreliable, subject to a number of biases and often unable to identify the most promising projects.

While the empirical evidence for most of these arguments is scattered, one result seems to be fairly robust. Multiple studies have shown that, on average, there is a declining marginal return on funding invested in research above a certain threshold. Although this threshold varies across disciplinary and national boundaries, it is not generally very high. However, reducing optimal funding to a question of evidence for or against concentration oversimplify a complex problem. The "proper" balance between concentration and dispersal of research funding is more a matter of degree: Both too little and too much concentration appears inefficient. Studies indicate that a healthy research system ecology includes both large and small groups (Wu et al., 2019). Furthermore, calls for greater diversity often seem to be interpreted as if all research can be perceived as equally good and equally important. This is obviously not the case, and if diversity in itself becomes a target, it may harm the overall development of science by diluting all quality concerns. As we highlight below, the need for a well-functioning quality assessment system may be even more pressing if more diversity should be pursued in a productive way, but it needs to build on a broader and more inclusive variety of academic and societal perspectives. Nonetheless, while the calls for diversity may be misused in certain situations, there are in general rather strong indications that most countries and fields need less, not more, funding concentration.

A number of potential remedies can be highlighted if the aim is to reduce further concentration: First, better oversight is needed within and across funding organizations to ensure that allocations are based on broader portfolio perspectives and less on assessments of individual applications in isolation. In particular, there is a need to monitor the success of the older and more established researchers, who appear to benefit the most from the lack of oversight. Second, experiments are needed with funding mechanisms seeking to counter the current concentration bias. A radical proposal here suggests widespread use of modified lottery models for grant applicants who pass an initial quality screening (Fang & Casadevall, 2016). Others suggest experimentation with new funding instruments to promote risky research—for instance by fully blinding the review process. These suggestions may, however, be at odds with aims to support more societally relevant research and should therefore only be a part of a broader portfolio of funding instruments. But most importantly, there is a need to start operating with broader understandings of research quality. Here it must be acknowledged more explicitly that "excellence" is multifaceted and multidimensional. Hence, allocation mechanisms must be better equipped to capture and reward the inherent variety of academic and societal dimensions, and must do so without ending up in a situation where the concept of quality disappears completely.

AUTHOR CONTRIBUTIONS

Emil Bargmann Madsen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. Kaare Aagaard: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

This work was supported by the Research Council of Norway, grant number 256223.

DATA AVAILABILITY

The data on individual research grants were obtained from individual public and private research funding organizations. Unfortunately, not all the funding organizations allow the data to be publicly available in a data repository.

REFERENCES

Aagaard, K. (2017). The evolution of a national research funding system: Transformative change through layering and displacement. *Minerva*, *55*(3), 279–297. https://doi.org/10.1007/s11024-017-9317-1

Aagaard, K., Nielsen, M. W., & Kladakis, A. (2020). Concentration or dispersal of research funding? *Quantitative Science Studies*, *1*(1), 117–149. https://doi.org/10.1162/qss_a_00002

Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*, *111*(16), 5773–5777. https://doi.org/10.1073/pnas.1404402111

Allison, P. D. (1980). Inequality and scientific productivity. *Social Studies of Science*, *10*(2), 163–179.

Allison, P. D., & Stewart, J. A. (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review*, *39*(4), 596–606.

Anand, S., & Hanson, K. (1997). Disability-adjusted life years: A critical review. *Journal of Health Economics*, *16*(6), 685–708. https://doi.org/10.1016/S0167-6296(97)00005-2

Best, R. K. (2012). Disease politics and medical research funding. *American Sociological Review*, *77*(5), 780–803. https://doi.org/10.1177/0003122412458509

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, *115*(19), 4887–4890. https://doi.org/10.1073/pnas.1719557115

Bowen, A., & Casadevall, A. (2015). Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. *Proceedings of the National Academy of Sciences*, *112*(36), 11335–11340. https://doi.org/10.1073/pnas.1504955112
Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687. https://doi.org/10.1038/nature18315
Cassi, L., Lahatte, A., Rafols, I., Sautier, P., & de Turckheim, É. (2017). Improving fitness: Mapping research priorities against societal needs on obesity. *Journal of Informetrics*, *11*(4), 1095–1113. https://doi.org/10.1016/j.joi.2017.09.010
Cole, S. (1970). Professional standing and the reception of scientific discoveries. *American Journal of Sociology*, *76*(2), 286–306.
Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, *214*(4523), 881–886.
Danish Council for Research and Innovation Policy. (2020). *Forskningsfinansiering i Danmark—hvad er der i vente?* Copenhagen.
Danish Government. (2006). *Progress, innovation and cohesion: Strategy for Denmark in the global economy*. Copenhagen.
Evans, J. A., Shim, J. M., & Ioannidis, J. P. A. (2014). Attention to local health burden and the global disparity of health research. *PLOS ONE*, *9*(4), 1–9. https://doi.org/10.1371/journal.pone.0090147
Fang, F. C., & Casadevall, A. (2016). Research funding: The case for a modified lottery. *MBio*, *7*(2), e00694-16. https://doi.org/10.1128/mBio.00694-16
Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science*, *13*(2), 285–305.
GBD 2017 DALYs and HALE Collaborators. (2018). Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1859–1922. https://doi.org/10.1016/S0140-6736(18)32335-3
Gillum, L. A., Gouveia, C., Dorsey, E. R., Pletcher, M., Mathers, C. D., Charles, E., & Johnston, S. C. (2011). NIH disease funding levels and burden of disease. *PLOS ONE*, *6*(2), e16837. https://doi.org/10.1371/journal.pone.0016837
Gläser, J. (2019). How can governance change research content? Linking science policy studies to the sociology of science. In D. Simon, S. Kuhlmann, J. Stamm, & W. Canzler (Eds.), *Handbook on science and public policy* (pp. 419–447). Cheltenham: Edward Elgar.
Gläser, J., & Laudel, G. (2016). Governing science. *European Journal of Sociology*, *57*(1), 117–168. https://doi.org/10.1017/S0003975616000047
Gläser, J., & Velarde, K. S. (2018). Changing funding arrangements and the production of scientific knowledge: Introduction to the special. *Minerva*, *56*(1), 1–10. https://doi.org/10.1007/s11024-018-9344-6
Gross, C., Anderson, G., & Powe, N. (1999). The relation between funding by the National Institutes of Health and the burden of disease. *New England Journal of Medicine*, *340*(24), 1881–1887. https://doi.org/10.1056/NEJM199906173402406
Head, M. G., Fitchett, J. R., Nageshwaran, V., Kumari, N., Hayward, A., & Atun, R. (2016). Research investments in global health: A systematic analysis of UK infectious disease research funding and global health metrics, 1997–2013. *EBioMedicine*, *3*, 180–190. https://doi.org/10.1016/j.ebiom.2015.12.016
Hegde, D., & Mowery, D. C. (2008). Politics and funding in the U.S. public biomedical R&D system. *Science*, *322*(December), 1797–1798.
Hegde, D., & Sampat, B. (2015). Can private money buy public science? Disease group lobbying and federal funding for biomedical research. *Management Science*, *61*(10), 2281–2298. https://doi.org/10.1287/mnsc.2014.2107
Jones, R., & Wilsdon, J. (2018). *The biomedical bubble: Why UK research and innovation needs a greater diversity of priorities, politics, places and people*. London: Nesta.
Katz, Y., & Matter, U. (2019). Metrics of inequality: The concentration of resources in the U.S. biomedical elite. *Science as Culture*, 1–28. https://doi.org/10.1080/09505431.2019.1694882
Klavans, R., & Boyack, K. W. (2017). Research portfolio analysis and topic prominence. *Journal of Informetrics*, *11*(4), 1158–1174. https://doi.org/10.1016/j.joi.2017.10.002
Larivière, V., Macaluso, B., Archambault, É., & Gingras, Y. (2010). Which scientific elites? On the concentration of research funds, publications and citations. *Research Evaluation*, *19*(1), 45–53. https://doi.org/10.3152/095820210X492495
Leisyte, L., & Dee, J. R. (2012). Understanding academic work in a changing institutional environment. In J. Smart & M. Paulsen (Eds.), *Higher education: Handbook of theory and research* (vol. 25, pp. 123–206). Dordrecht: Springer.
Lerchenmueller, M. J., & Sorenson, O. (2018). The gender gap in early career transitions in the life sciences. *Research Policy*, *47*(6), 1007–1017. https://doi.org/10.1016/j.respol.2018.02.009
Li, D., & Agha, L. (2015). Big names or big ideas: Do peer-review panels select the best science proposals? *Science*, *348*(6233), 434–438. https://doi.org/10.1126/science.aaa0185
Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(10), 317–323.
Ma, A., Mondragón, R. J., & Latora, V. (2015). Anatomy of funded research in science. *Proceedings of the National Academy of Sciences*, *112*(48), 14760–14765. https://doi.org/10.1073/pnas.1513651112
Manton, K. G., Gu, X.-L., Lowrimore, G., Ullian, A., & Tolley, H. D. (2009). NIH funding trajectories and their correlations with US health dynamics from 1950 to 2004. *Proceedings of the National Academy of Sciences*, *106*(27), 10981–10986. https://doi.org/10.1073/pnas.0905104106
Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.
Mongeon, P., Brodeur, C., Beaudry, C., & Lariviere, V. (2016). Concentration of research funding leads to decreasing marginal returns. *Research Evaluation*, *25*(4), 396–404. https://doi.org/10.1093/reseval/rvw007
Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5
Myers, K. (n.d.). The Elasticity of Science. *American Economic Journal: Applied Economics*, 1–38.
Parks, R. (2014). The rise, critique and persistence of the DALY in global health. *The Journal of Global Health*, *4*(1), 28–32.
Peifer, M. (2017). The argument for diversifying the NIH grant portfolio. *Molecular Biology of the Cell*, *28*(22), 2935–2940. https://doi.org/10.1091/mbc.E17-07-0462
Price, D. J. D. S. (1963). *Little science, big science*. New York, NY: Columbia University Press.
Reskin, B. F. (1977). Scientific productivity and the reward structure of science. *American Sociological Review*, *42*(3), 491–504.
Rygård, S. L., Kjær, M.-B. N., & Perner, A. (2018). Statens investering i kliniske forsøg. *Ugeskrift for Læger*, *180*(16), 2–4.
Sörlin, S. (2007). Funding diversity: Performance-based funding regimes as drivers of differentiation in higher education systems.

Higher Education Policy, *20*(4), 413–440. https://doi.org/10.1057/palgrave.hep.8300165
Stoeger, T., Gerlach, M., Morimoto, R. I., & Nunes Amaral, L. A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. *PLOS Biology*, *16*(9), 1–25. https://doi.org/10.1371/journal.pbio.2006643
UFM. (2016). *Private Fonde: En kortlægning af bidraget til dansk forskning, innovation og videregående uddannelse*. Copenhagen.
UK Clinical Research Collaboration. (2015). *UK Health Research Analysis 2014*. London.
Wahls, W. P. (2018a). High cost of bias: Diminishing marginal returns on NIH grant funding to institutions. *BioRxiv*. https://doi.org/10.1101/367847
Wahls, W. P. (2018b). The NIH must reduce disparities in funding to maximize its return on investments from taxpayers. *ELife*, *7*, 1–9. https://doi.org/10.7554/eLife.34965
Wang, J., Lee, Y. N., & Walsh, J. P. (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. *Research Policy*, *47*(6), 1070–1083. https://doi.org/10.1016/j.respol.2018.03.014
Whitley, R., Gläser, J., & Laudel, G. (2018). The impact of changing funding and authority relationships on scientific innovations. *Minerva*, *56*(1), 109–134. https://doi.org/10.1007/s11024-018-9343-7
Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*, 378–382. https://doi.org/10.1038/s41586-019-0941-9
Yao, L., Li, Y., Ghosh, S., Evans, J. A., & Rzhetsky, A. (2015). Health ROI as a measure of misalignment of biomedical needs and resources. *Nature Biotechnology*, *33*(8), 807–811. https://doi.org/10.1038/nbt.3276
Yegros-Yegros, A., Klippe, W. Van De, Abad-Garcia, M.-F., & Rafols, I. (2019). Exploring why global health needs are unmet by public research efforts: The potential influences of geography, industry, and publication incentives. https://doi.org/10.2139/ssrn.3459230
Zuckerman, H. (1970). Stratification in American science. *Sociological Inquiry*, *40*(2), 235–257.

APPENDIX

Table A1. Summary statistics for full and Web of Science-matched samples

	Full sample	WoS-matched sample
Female grantees	29%	27%
Funding amount		
Mean	2,472,936	2,650,834
Median	934,001	1,000,000
Grant type		
Research project	56,2%	59,1
Postdoc grant	12, 6%	10,9%
Studentship grants	7,1%	7,5%
Mobility grants	5,8%	5,3%
Infrastructure grants	4,2%	5,2%

Table A2. Percentiles of funding for full and Web of Science-matched samples

Percentiles	Full sample	WoS-matched sample
10th	120,000	133,798.4
20	200,000	221,702.4
30	350,000	400,000
40	513,399	600,000
50	970,076	1,000,000
60	1,367,621	1,465,663.8
70	1,824,608	1,977,948.3
80	2,586,948	2,750,013.8
90th	5,573,906	5,796,784.9

Table A3. Summary of cumulative funding amounts by academic age

Academic age	Mean amount	Median amount	Std. dev.	25th percentile	75th percentile	*N*
1	2,059,150	1,394,369	3,982,689	465,050	2,272,460	203
2	1,916,986	1,406,368	2,838,634	491,842.2	2,216,921	274
3	2,359,436	1,281,250	4,524,869	612,500	2,319,287	289
4	2,318,330	1,430,635	5,613,875	475,133	2,639,520	365
5	3,560,705	1,471,361	8,873,202	457,283.5	2,920,518	360
6	3,919,500	1,811,427	6,725,612	700,000	3,692,522	287
7	5,212,371	1,975,000	8,913,655	678,674	5,845,278	277
8	5,816,947	2,587,626	9,279,674	759,628	6,367,089	298
9	8,104,687	3,548,362	12,941,586	1,288,375.5	10,807,996	204
10	9,013,589	4,231,821	13,069,892	1,995,207.8	10,319,452	224
11	10,519,679	5,845,956	16,244,098	2,048,200	12,443,770	206
12	11,831,184	4,406,880	27,560,715	1,416,200	10,331,734	247
13	16,941,984	6,237,916	30,578,227	2,464,663.8	15,520,343	214
14	12,040,523	6,037,999	17,240,231	2,274,275	14,029,316	217
15	18,532,573	8,334,335	28,180,229	3,959,349	20,939,255	233
16+	20,350,703	10,503,805	28,182,465	4,257,937	20,436,796	150

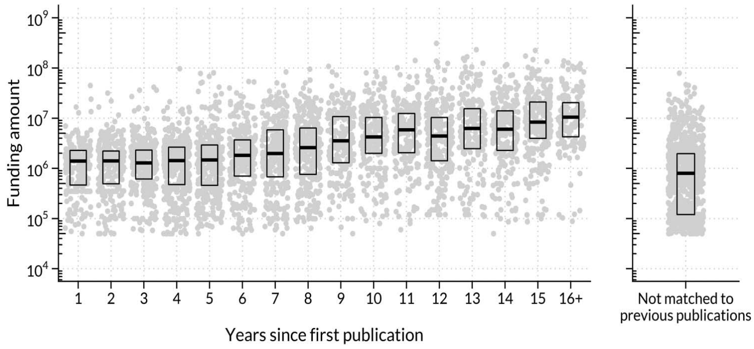


Figure A1. Distribution of cumulative funding across academic age. Boxplots show 25th, 50th, and 75th percentiles.

9. Thematic funding and topic switching in a sample of 10,475 scientists

Emil Bargmann Madsen[1] and Mathias Wullum Nielsen[2]

Thematic funding instruments have been criticised for restricting researchers' creativity by limiting the topics they can obtain funding for and incentivising topic shifts. However, the empirical evidence backing such claims is scarce. We use citation-based community detection to examine topic changes in a United Kingdom-based sample of 10,475 recipients of targeted and non-targeted research grants in Physics, Engineering, and Bioscience. Using coarsened exact matching and a cosine similarity measure, we find that researchers that acquire targeted funding are more likely to change their topic focus than are comparable researchers funded through non-targeted schemes. However, we also find that recipients of both types of grants tend to revert toward their original research areas when the funding expires. We find no evidence that targeted funding increases the topic diversity in a recipient's research portfolio. Our findings highlight the difficulties of steering research content by means of thematic funding.

[1]Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University

[2]Department of Sociology, University of Copenhagen

9.1 Introduction

A scientist's choice of research problem is a multifaceted process involving a number of considerations (Gieryn, 1978; Zuckerman, 1978). Scientists are often commanding a significant amount of leeway over what topics to research, while also being constrained or enabled by the priorities set by sponsors (Evans, 2010; Franssen et al., 2018; Luukkonen & Thomas, 2016; Ziman, 1987). Faced with decreasing budgets, public research agencies are becoming increasingly strategic in their priorities (Klavans & Boyack, 2017a). Science is expected to contribute to the national economy, and funders use targeted instruments to exercise authority over the types of research conducted (Brattström & Hellström, 2019; Gläser & Laudel, 2016; Hellström et al., 2017; Stewart, 1995). One such instrument is thematic funding that channels resources into prioritised research areas. Thematic funding has been criticised for restricting researchers' creativity by limiting the topics they can obtain funding for and incentivising topic shifts (Nature, 2003; Nuffield Council on Bioethics, 2014). However, the empirical evidence supporting such claims is sparse. In this paper, we examine the relationship between thematic funding priorities and scientists' topic choices.

The question of how thematic priorities influence topic choices is important for at least two reasons. First, thematic funding may influence the 'essential tension' between tradition and change in science (Kuhn, 1977). To be successful, a scientist is bound to conduct research within the tradition of her research field, while also breaking free of the scientific consensus to make new discoveries (Bourdieu, 1975; Foster et al., 2015; Kuhn, 1977). Thematic funding can incentivise scientists to focus on well-established topics (Foster et al., 2015; Stoeger et al., 2018) and this may slow the pace of scientific progress (Rzhetsky et al., 2015). Second, thematic funding may cause undue skewness in what topics are investigated and "dry out" areas of modest political interest (Gläser & Laudel, 2016).

Quantitative research shows that scientists generally limit their focus to a few topics, and that topic switching is relatively rare. However, topic switching has become more prominent over the years and the popularity of research topics seems to coincide with government priorities, individual career developments, and historical events (Aleta et al., 2019; Domenico et al., 2016; Horlings & Gurney, 2013; Jia et al., 2017; Zeng et al., 2019). Qualitative studies suggest that researchers are attentive to funding priorities and tailor their grant proposals to match prioritized areas (Gläser & Laudel, 2016; Laudel, 2006a, 2006b; Laudel & Gläser, 2014). So far, only one study has quantified how scientists respond to thematic funding priorities. Based on data from 435 thematic grant schemes released by the US National Institutes of Health, the study shows that most recipients of thematic grants are already active in research areas close to the prioritized topics.

Moreover, many of these recipients tend to return to their original research areas when the thematic grants expire (Myers, 2020).

We add to this literature, by conducting the first quasi-experimental study aimed at quantifying the extent to which thematic funding induce United Kingdom (UK) scientists in Engineering, Physical Sciences, and Bioscience to switch topics. Our study sheds light on two possible ways in which thematic funding may influence topic selection and resource allocation.

- Thematic funding priorities may induce topic shifts through a change in the funding system's incentives. Research suggests that shifts in topic focus are often associated with important career moves, such as transitioning to a Postdoc or getting tenure (Horlings & Gurney, 2013). Thematic funding could credibly induce similar shifts. In qualitative interviews, researchers report having adapted to funding programs by selecting externally predetermined topics (Laudel, 2006b; Laudel & Gläser, 2014; Luukkonen & Thomas, 2016). However, such topic shifts may only be temporary, as grant recipients tend to return to their original focus areas when the thematic funding expires (Myers, 2020).
- Thematic priorities may induce a cross-fertilisation between researchers' existing topic foci and new topic branches. For instance, qualitative research suggests that some scientists formulate proposals to bridge their own topic interests with that of research sponsors (Luukkonen & Thomas, 2016).

To examine how funding priorities influence topic selection, we build a database of 10,475 grants awarded to researchers in Engineering, Physical Sciences, and Bioscience from Research Councils UK, and partition grantees into recipients of either non-targeted funding or targeted funding. We extract publication data for all grantees in Clarivate's Web of Science (WoS) and create individual time-series of research topics using citation-based community detection. We use these topics to determine whether the grantees' research focus changes after they receive a targeted or non-targeted grant. We use coarsened-exact matching to examine variations in topic distributions for the two groups of recipients before and after receiving a grant. Specifically, we partition each PI's publications into three distinct and non-overlapping period groups: The 'before' group represents publications published before a grant period begins, the 'during' group refers to publications that are the direct outcome of a grant, and the 'after' group comprise articles published after the grant has expired. For the papers published in each period group, we record the distribution of topics, and compare these distributions using a cosine similarity measure. Moreover, we use the Rao-Stirling diversity index to discern whether targeted grants induce cross-fertilisation by diversifying researchers' topic portfolios (see methods for a breakdown of the full procedure). In the following, we use the terms thematic funding/targeted funding/priority funding interchangeably.

9.2 Results

Figure 9.1 reports the topic-similarity distributions for recipients of non-priority and priority grants across the three period groups. A similarity score of zero indicates no overlap across period groups; a score of one suggests perfect overlap. As indicated by the high, average similarity scores across the six distributions, both recipients of priority and non-priority grants appear to study topics related to their prior work. This finding aligns with recent research suggesting that many recipients of priority grants may already be active in research areas close to the prioritised topics (Gläser, 2019). However, PIs with priority grants have topic portfolios that differ more from their past publications than do PIs with non-priority grants, with mean similarities of 0.57 vs. 0.66 (diff. = 0.09) and median similarities of 0.66 and 0.78 (diff. = 0.12) for the comparison of publications published before and during the grant period.

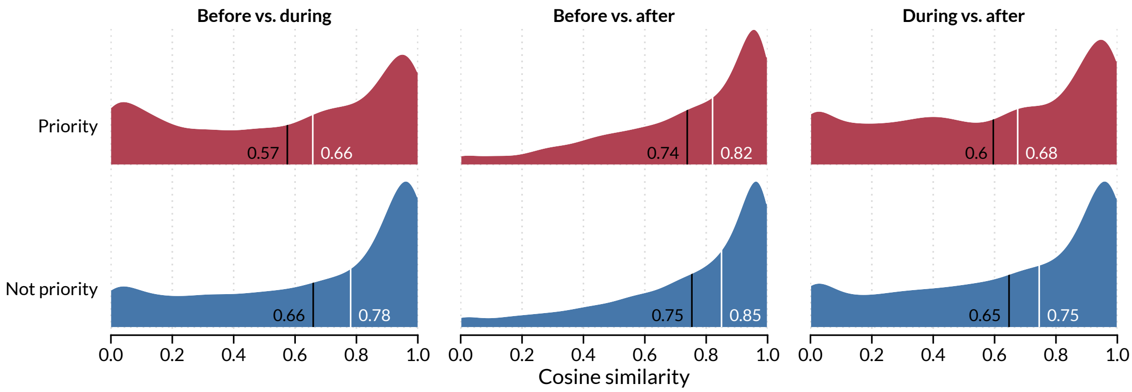


Figure 9.1: Topic-similarity distributions for priority and non-priority grants. Vertical black lines specify within-group mean similarities. Vertical white lines specify within-group medians. N = 10,475.

After the grant period, both types of grant holders revert toward their previous research areas, with mean similarities of 0.74 and 0.75 (diff. = 0.01) and median similarities of 0.82 and 0.85 (diff. = 0.03) for the comparison of papers published *before* and *after* the grant period. Lastly, priority grant holders have marginally lower similarity scores than do non-priority grant holders for the comparison of papers published during and after the grant period (mean similarity = 0.60 vs. 0.65, diff. = 0.05; median similarity = 0.68 vs. 0.75, diff = 0.07).

In sum, these descriptive findings suggest that priority grants induce larger topic shifts than do non-priority grants, although this effect tends to revert as the publication activities related to a grant level out. The observed variation in similarities for recipients of non-priority and priority grants is most salient on the left side of the distributions in figure 9.1. When comparing papers published before and during a grant period, 24.2 % of PIs with non-priority grants have similarities in the lowest 25th percentile (i.e. most different), while 33.2 % of the PIs with priority grants have similarities within this percentile range (see Supplementary figure S9.3). When comparing papers published before and after a grant period, these numbers are 24.8 % for the non-priority group and 27.7 % for the priority group.

However, topic changes are multifaceted, and prior research suggests that a change in focus is most likely to occur (i) when a researcher transfers from being a PhD student to being a Postdoctoral researcher (Horlings & Gurney, 2013), (ii) at later stages in the academic career (Zeng et al., 2019), and (iii) among researchers with low citation impact (22). To account for these possible confounders, we match the recipients of priority and non-priority grants on seven background variables, using Coarsened Exact Matching (See Methods for full details).

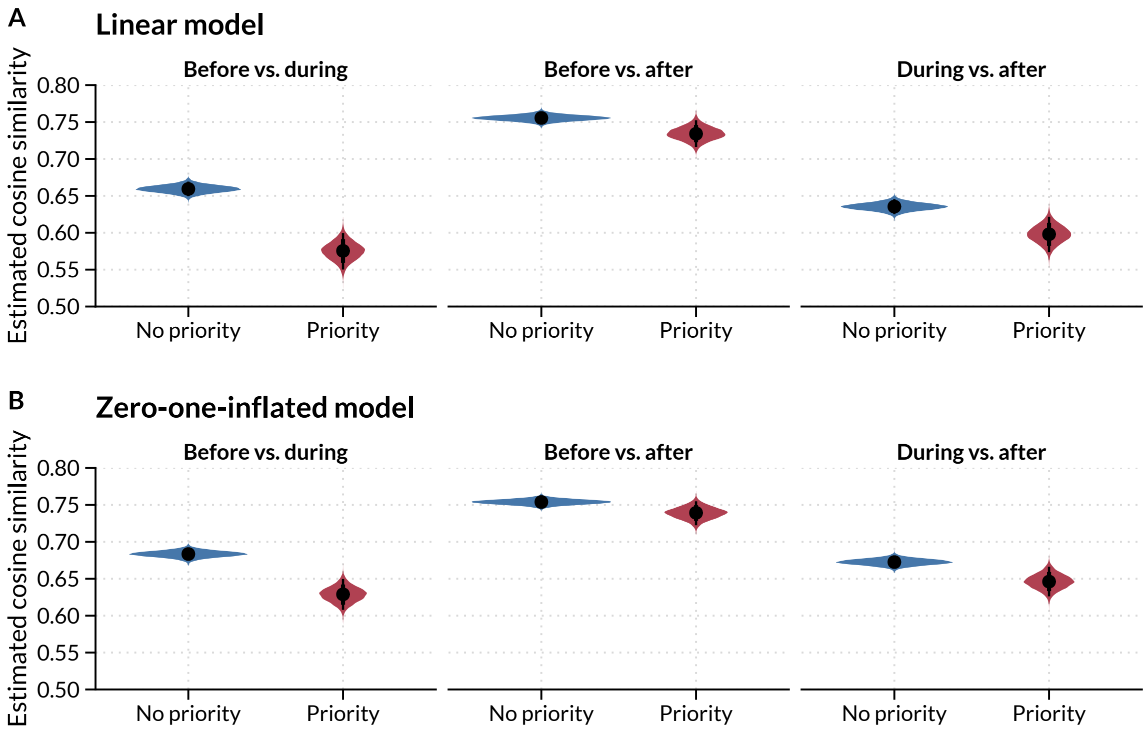


Figure 9.2: Estimated topic similarities for researchers receiving prioritised or non-prioritised funding. (A) Estimated mean similarities derived from the coarsened exact matched sample using a Bayesian linear model. (B) Estimated mean similarities derived from the coarsened exact matched sample using a Bayesian zero-one-inflated beta model. Point estimates are medians of the posterior distributions (coloured violin plots), with 80 % (thick lines) and 95 % (thin lines) credible intervals. N = 5,184

Figure 9.2 reports the estimated mean similarity scores derived from the matched sample, using a linear model (A), and a generalized linear model with a zero-one-inflated beta distribution (B). First, we report the outcomes of the linear model. Congruent with the descriptive evidence presented above, we find that recipients of priority grants, on average, alter their topic focus more than do recipients of non-priority grants. Comparing the similarity of papers published before and during a grant (Figure 9.2, panel a), we observe a difference in estimated means of -0.084 (95 % CI: [-0.111 to -0.0581]) between recipients of priority and non-priority grants, which is equivalent to a Cohen's d = -0.243.

When comparing the similarity of research published before and after the grant period, we observe a notably smaller difference between the priority and non-priority groups (estimated mean difference= -0.021, 95 % CI: [-0.041 to -0.002], Cohen's d = 0.082). The matched analysis thus corroborates the descriptive result that grantees, irrespective of grant type, revert toward their previous research area when their grants expire.

Lastly, the comparison of papers published during and after a grant shows a difference in estimated mean similarity of -0.037 (95 % CI: [-0.063 to -0.012], Cohen's d = -0.111), between recipients of priority and non-priority grants, with priority grantees scoring slightly lower on the similarity score than non-priority grantees (Figure 9.2). In conjunction, these findings suggest that both priority and non-priority grantees combine grant-specific topics with earlier research interests in the publication period exceeding their grants.

These results largely correspond with the outcomes of the zero-one-inflated beta (ZOIB) model (Figure 9.2, panel B), albeit the estimated mean differences are smaller in the latter model. Using the ZOIB model to estimate the mean similarity scores for papers published by recipients of priority and non-priority grants before and during the grant period, we observe a difference of -0.055 (95 % CI: [-0.077 to -0.032], Cohen's d = -0.06). The ZOIB model also estimates the presence of extreme similarity values in both tails of the distribution, and shows that the general probability of scoring either 0 or 1 is only 0.02 (95 % CI: [-0.002 to 0.044]), or 2%, in the sample. However, given an extreme similarity of either 0 or 1, recipients of priority grants have a 21 % lower probability than do recipients of non-priority grants of studying exactly the same topics as previously studied (non-priority grantees, conditional probability: 0.33, 95 % CI: [0.284 to 0.377]; priority grantee, conditional probability: 0.12, 95 % CI: [0.066 to 0.207]). For the remaining two comparisons, the zero-one-inflated model mirrors the linear model closely. The before vs. after similarities are high for both grant types (mean diff. = -0.015, 95 % CI: [-0.033 to 0.002], Cohen's d = -0.023), and the during vs. after similarities differ only by -0.026 (95 % CI: [-0.048 to -0.005], Cohen's d = -0.032).

To verify the robustness of these results, we estimated 72 regressions with different model specifications and data choices to assess the variation in effect sizes across the three outcome measures (see Methods). Figure 9.3 shows the difference in estimated means across 36 linear models and 36 ZOIB models. These robustness checks generally confirm the findings presented above. When comparing papers published before and during a grant period, priority grantees tend to change their topic focus more than non-priority grantees (Figure 9.3). The average difference across models is -0.054. For the comparisons of papers published before and after a grant period and during and after a grant period, the differences in mean similarities are smaller, with most or some of the estimates being close to zero.

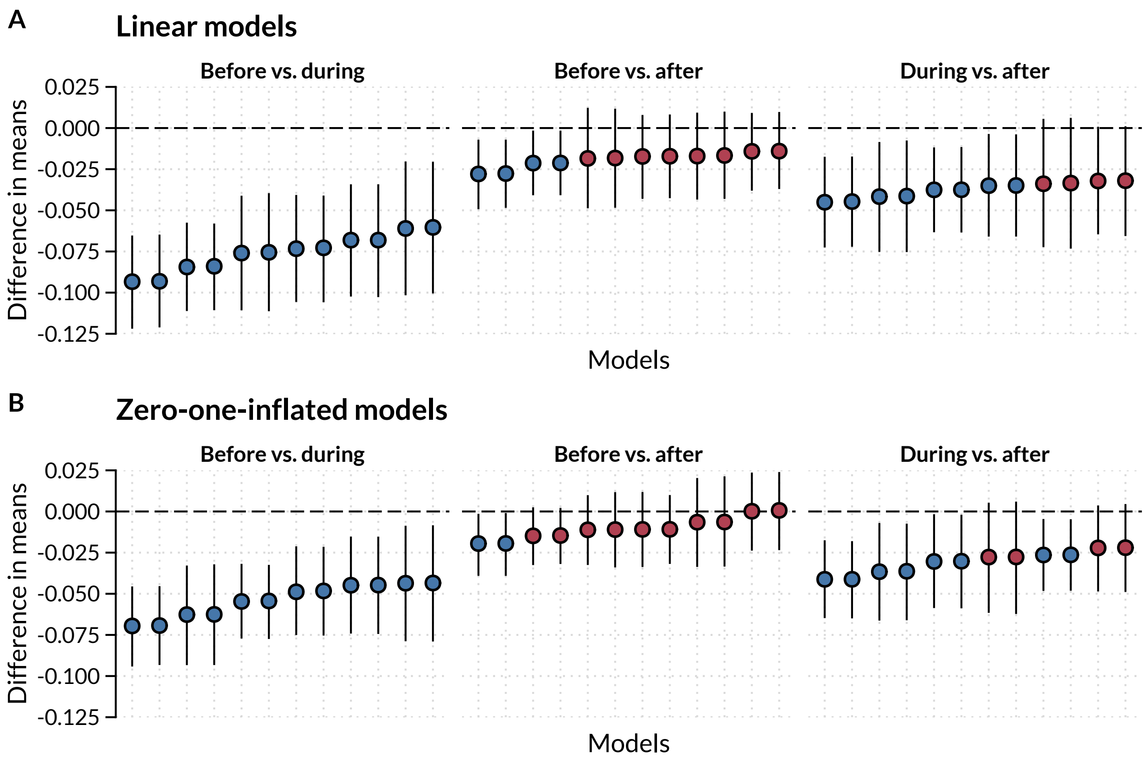


Figure 9.3: Estimated differences in topic similarities. (A) Estimated mean difference in similarities from Bayesian linear models. (B) Estimated mean difference in similarities from Bayesian zero-one-inflated beta models. Point estimates are medians of posterior distributions with 95 % credible intervals.

As a final step in the analysis, we examine whether priority grants induce scientists to bridge existing research activities with new focus areas, thereby diversifying their portfolios. For each grantee type and period, we use citation data to calculate the Rao-Stirling diversity index, which is a composite measure of how many different topics a researcher studies, how much of a research portfolio each topic takes up, and how closely the topics relate through citations (see Methods).

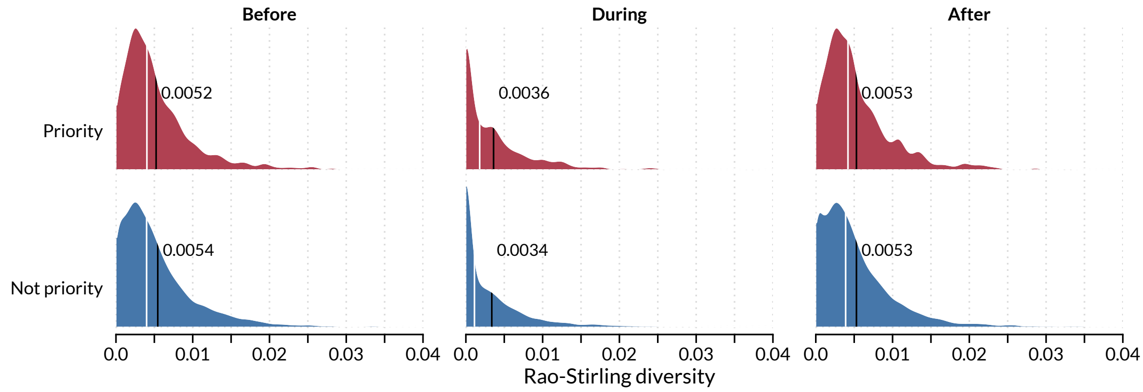


Figure 9.4: Distributions of topic diversity for priority and non-priority grants. Vertical black lines are within group mean diversity, while vertical white lines are within group medians. N = 10,475.

Figure 9.4 reports the distribution of topic-diversity scores for each grantee type and period. Generally, grant recipients tend to work on topics that are closely related, and we find no evidence that the topic portfolios of priority grantees diversify more after receiving a grant than do the topic portfolios of non-priority grantees. Both types of recipients work on a less diverse portfolio of topics during the grant period, which is expectable given the limited scope of a single grant. As in our analysis of topic similarity, we match priority and non-priority grants and compare these using a linear random intercept model. Figure 9.5 shows the posterior distributions and medians. The results confirm that grantees generally work on closely related topics, and that their work only diversifies after a grant period, albeit to a small extent. Moreover, priority and non-priority grant holders do not differ markedly. The estimated difference in diversity is 0.0000996 (95 % CI: [-0.000218 to 0.000406]) before the grant, 0.000121 (95 % CI: [-0.000196 to 0.000430]) during the grant, and 0.000208 (95 % CI: [-0.000113 to 0.000543]) after the grant.

9.3 Discussion

Research agencies are increasingly strategic in their funding activities. Thematic priorities have become an important lever for steering scientists toward targeted research areas (Gläser, 2019). Agencies set priorities, and expect scientists to follow the money (Myers, 2020). However, the effects of thematic priorities on scientists' topic choices remain uncertain. To what extent are scientists adapting to these priorities by navigating their research programmes in new directions; and do thematic funding promote diversity and cross-fertilization by inducing scientists to bridge their existing interests with those of research sponsors?

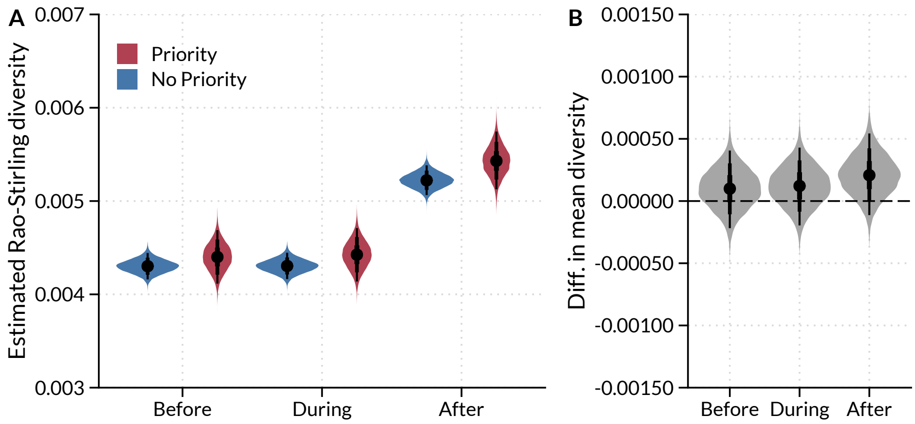


Figure 9.5: Estimated topic diversity. (A) Estimated mean diversity derived from the coarsened exact matched sample using a Bayesian linear random intercept model. (B) Estimated differences in mean diversity derived from the coarsened exact matched sample using a Bayesian linear random intercept model. Point estimates are medians of the posterior distributions (violin plots), with 80 % (thick lines) and 95 % (thin lines) credible intervals. N = 5,184. The model-based means are estimated from a smaller sample, as not all non-priority grants were matched.

In this study, we have attempted to answer these questions by gauging trends in topic selection in a UK-based sample of 10,475 recipients of targeted and non-targeted research grants in Physics, Engineering, and Bioscience.

Our analysis suggests that grantees generally study topics related to their prior work. However, recipients of targeted grants, on average, alter their topic focus more than do recipients of non-targeted grants, although this difference is relatively small and levels off when the grant period ends. In addition, we observe near-identical levels of topic diversity in the research portfolios of targeted and non-targeted grant recipients before, during and after a grant. This suggests that targeted grants do not induce higher levels of cross-fertilization between distant topics than do non-targeted grants.

Our study highlights the difficulties of steering research by means of thematic funding mechanisms. Agencies typically use such mechanisms to intensify the research capacity and capability in areas of high political priority. However, if most grantees are either already active in a prioritized area or tend to revert to their prior focus when the grant funding expires, thematic funding may fail to produce the desired outcomes.

The low levels of topic diversity observed in this study align with evidence suggesting disproportionate success rates for interdisciplinary funding proposals (Bromham et al., 2016). One driving factor may be that grant reviewers perceive such proposals as too risky and downgrade them as less "doable" (Bromham et al., 2016). Another explanation may lie in the procedural aspects of grant reviewing, where it is not unusual to rank applicants on their past performance. In such scenarios, scientists' concerns about "staying competitive" may crowd out motivations to engage in projects that span distant topics, where "switching costs" (including cognitive and collaborative challenges) will lower productivity-levels in the short term (Leisyte et al., 2010; Myers, 2020).

Our study has four limitations. First, despite a carefully designed matching strategy, unmeasured confounding may possibly bias our analysis. Second, our study does not include a control group of researchers without funding, which prevents us from concluding anything about how recipients of targeted grants differ from the "ordinary scientist" with respect to topic switching. Third, the increasing concentration of research funding in the hands of a select group of highly successful PIs (Hand & Wadman, 2008; Katz & Matter, 2020; Mongeon et al., 2016) implies that some grantees may be leading several funded projects simultaneously. This makes it difficult fully parcel out the association between targeted grants and topic switching. Fourth, the strength of the association between targeted grants and topic switching evidently depends on the specificity of the prioritized research areas, as broader topics make it easier for scientists to match their existing research activities to thematic priorities (Laudel, 2006b). In this study, we have focused on the priorities set by two UK research councils, and future research may reach different conclusions for other councils.

Our study also has four main strengths. First, building on all grants funded through two major UK research councils over eight years, our study provides one of the first large-scale assessments of how targeted grants affect scientists' problem choices. Second, our comparison of topic similarities across papers published before, during and after a grant, allow us to estimate both the short-term and longer-term impact of targeted funding on topic switching and diversity. Third, while not perfect, our matching of PIs on pre-grant citation impact and publication rates, academic age, discipline, funding amount, type of grant, and funding organization, has likely eliminated important confounding in the cross-group comparisons. Fourth, our findings are robust across a large set of varying model specifications and data choices. This increases the transparency and reliability of our conclusions.

In sum, we establish an empirical link between targeted funding and topic switching. Our findings suggest that while thematic funding induces scientists to shift their topic focus in the short term, the longer-term effects are negligible. Our study serves as a reminder that many factors in addition to funding affect scientists' problem choices, and that radical shifts in research interests are relatively rare.

9.4 Methods

9.4.1 Data and context

We used data from the U. K. *Engineering and Physical Sciences Research Council* (EPSRC) and the *Biotechnology and Biological Sciences Research Council* (BBSRC). Together, these research councils invest £1.3 billion annually in a diverse set of research disciplines within the natural sciences, covering fields such as healthcare technology, engineering, materials science, molecular biology, chemistry and mathematics. From the EPSRC's Grants on the Web and BBSRC's advanced grant search, we collected meta-data on the principal investigator (PI), affiliation, grant amounts, grant type, start- and end dates, and grant abstracts for an initial sample of 67,309 research grants.

The EPSRC and the BBSRC primarily fund researcher-initiated projects, but also make funding available for projects within specific, prioritised areas. In the BBSRC, the priority areas cover broad strategic topics (e.g. combatting anti-microbial resistance) that the funding council wishes to promote or encourage, and while all proposals are subject to the same assessment criteria, a project's relevance to the strategic topic areas is included as a factor in the appraisal process (see Supplementary figure S9.1 and S9.2 for differences in success rates). In the ESPRC, the success rates have consistently been higher for targeted proposals than for non-targeted proposals.

For the EPSRC, we focused on grants funded before 2012, when the council distinguished between grants in responsive mode (non-targeted) and targeted mode. We classified each grant by its mode as reported in the council's web database of grants, yielding 1,499 priority grants. For the BBSRC, we recorded 683 grants that addressed one or more of the council's strategic priority areas as targeted, and denoted the remaining BBSRC grants as non-targeted.

We linked each PI to her publication history through multiple steps. First, we matched the DOI of publications reported in the UKRI Gateway to Research to publications indexed in WoS, and then matched to PI names via the relative Levensthein distance. To increase the matching rate, we also created a set of likely e-mail addresses in the formats 'firstname.lastname@institution.ac.uk' or 'initial.lastname@institution.ac.uk'. In total, the matching process resulted in a base sample of 31,839 grants and 609,709 publications, spanning both the period prior to a grant ('before'), publications resulting from a grant ('during'), and publications published after a grant that are not related to the grant. All publications with the PI as an author, but not reported as an output from the grant in question, was deemed not related to a grant.

To ensure comparability across grants, we constructed three final samples using the following criteria: In sample A, each grantee should have a minimum of six publications with two publications in each of the periods (before, during, and after). For sample B, each grantee should have at least nine publications, with three in each period. For sample C, each grantee should have at least 15 publications, with five in each period. Because many grantees have few publications, the three samples include fewer grants as the required number of publications increase 9.1. All main analyses were based on sample A, while robustness tests in 9.3 also relied on samples B and C.

Table 9.1: Grant samples.

	Non-priority grants	Priority grants	Total grants
Sample A	9,575	900	10,475
Sample B	7,758	739	8,497
Sample C	5,012	489	5,501

9.4.2 Measuring topic change and diversity

To determine the distribution of topics in a scientist's research portfolio, we follow Klavans and Boyack (2017c) and Waltman and van Eck (2012) and draw each publication's topic from a global map of science produced by the Centre for Science and Technology Studies (CWTS), Leiden University. The CWTS topic map is based on the weighted direct citation network of 20 million publications in the Web of Science 2000-2017, clustered using the Smart Local Moving Algorithm. Papers are assigned to a topic cluster based on their citing, citation and topical commonalities. These clusters thus define small fields with common referencing cultures, and topics can be thought of as a cognitive space populated by documents occupied with common problems or at least referring to shared problems (Klavans & Boyack, 2017b). Topic categorisations can be based on different granularities depending on how specific the topics should be. Our analysis relies on a categorisation with 868 topics, but we recreate the key findings using a more fine-grained categorisation with 3,972 topics (see Supplementary S9.4).

Our study investigates how funding arrangements affect both the diversity of a scientist's research portfolio and the change in this portfolio. While these factors are inextricably linked, they tell separate tales about a scientist's career trajectory. Diversity is a question of the variety, balance, and similarity of the topics considered by a scientist, while change concerns the magnitude and direction of shifts between topics.

We operationalise a scientist's research portfolio as a distribution of publications across topic clusters, and compute individual distributions of topics for each of the analytical periods: Before a grant, during a grant, and after a grant. First, we investigate how similar or dissimilar the topics investigated by a scientist are before, during and after a grant event. Our approach measures changes in a PIs topic distributions by comparing the distribution of topics over all publications before a grant, with the distribution of topics from publications produced as part of the grant project, and the publications produced after a grant. For each of these three distributions, topics are gathered in a topic vector P, with author i (the scientist) as the sole row, and all 868 possible topics as columns. Each vector cell scores the relative proportion of papers published on a given topic. For each scientist, we compare all combinations of the three topic-distribution vectors, by computing the cosine similarity for two vectors P and Q:

$$S(P,Q) = \frac{\sum_{i=1}^{n} P_i \cdot Q_i}{\sqrt{\sum_{i=1}^{n} P_i^2} \cdot \sqrt{\sum_{i=1}^{n} Q_i^2}}$$

This yields a symmetric 3 × 3 matrix of similarities for each scientist, and three similarity measures. The similarities take values between 0 (= completely different) and 1 (= completely similar):

	Before	During	After
Before	1	S_1	S_2
During	S_1	1	S_3
After	S_2	S_3	1

In supplementary figure S9.5 and S9.6, we show that using the Euclidean distance instead of the cosine similarity do not substantially alter our results.

Second, we measure changes in topic diversity by comparing the Rao-Stirling diversity index (Stirling, 2007) for the individual scientist before, during and after a grant period. The Rao-Stirling index accounts for both the variety of topics, the balance (the proportion of the portfolio they take-up) and disparity (how similar they are in the first place) between them. For each year, we calculate the index as

$$D = \sum_{ij(i \neq j)} p_i \cdot p_j \cdot d_{ij}$$

with p_i and p_j being the proportions of articles going to topic i and j, and d_{ij} being the disparity between the topics. To measure disparity, we calculate the cosine similarity between topics based on how often a paper on one topic cites a paper on another topic.

9.4.3 Comparing targeted and non-targeted grants using matching

Selection effects and confounding may bias direct comparisons of how targeted and non-targeted funding schemes influence grant recipients' topic choices. To reduce possible confounding, we used coarsened exact matching (Iacus et al., 2011, 2012) of the two PI groups. The matching of the two types of grant holders will likely not eliminate unmeasured confounding, but can serve as a method of pre-processing the data to reduce bias, model dependency, and ensure more comparable groups of grantees (Ho et al., 2007).

Specifically, we created strata of similar PIs, and compared topic diversity and change within these strata. Matches between recipients of targeted and non-targeted funding were made on a set of individual and grant-specific attributes prior to the time a PI received her grant. We matched on the following attributes: (i) total field-normalised citation scores (*TNCS*), (ii) number of pre-grant publications, (iii) academic age measured from first published paper to the grant year, (iv) most prevalent academic discipline, (v) funding amount, (vi) type of grant (research grant, fellowship or other), and (vii) funding organization (EPSRC or BBSRC). Section B in the supplementary material gives an overview of the co-variate balance between the unmatched and matched groups.

The matching procedure followed a combination of coarsened exact matching (CEM) on attribute (i)-(v) and exact matching on attribute (vi-vii). The coarsening of the attributes and the subsequent matching were done by automatically binning (i)-(iv) and exactly matching the two groups on the binned values of each variable using the R program cem (Iacus et al., 2009). The matching procedure creates a number of stratums, and compares observations within these stratums. For the analysis, we created a normalized weight W with $W = 1$ for a scientist with a targeted grant:

$$W = \frac{m_C}{m_T} W_S$$

for a scientist with a responsive mode grant, and $W = 0$ for non-matched scientists. Here, m_C and m_T are the number of control units and treated units in the whole sample, and

$$W_s = \frac{m_C^S}{m_T^S}$$

where S denotes the matching stratum. As a robustness check, we also matched the two PI groups using Inverse-Probability-Weighting. Weights were generated by estimating a logistic regression with matching attributes (i)-(viii) on an indicator variable P_i of 0 for non-targeted grants and 1 for targeted grants. For each grant we calculated a predicted propensity e_i of being a targeted grant, and created a weight as

$$W = \frac{e_i P_i}{e_i} + \frac{e_i(1 - P_i)}{1 - e_i}$$

This yields an average treatment effect of the treated by keeping the treatment population, i.e. the priority grants, and weighting the control group, the non-priority grants, to match them.

9.4.4 Models used for group comparisons

To investigate whether targeted grants induce larger topic changes and more diversity in a researcher's portfolio, we estimate group differences for four different measures. First, we compare the three similarity measures (before vs. during, before vs. after, and during vs. after) between non-priority and priority grants via a simple Bayesian weighted linear regression, with:

$$S \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 P_i + \epsilon$$
$$\sigma = \gamma_0 + \gamma_1 P_i + \epsilon$$

where β_0 gives the mean similarity for the group of non-priority grants, $\beta_0 + \beta_1$ is the mean similarity for the group of priority grants, and ϵ is the residual variance. Note that because the groups of priority and non-priority grantees are of very different sizes, we assume an unequal variance, and model this variance. Thus for σ, γ_0 gives the standard deviation for non-priority grantees, and $\gamma_0 + \gamma_1$ gives the standard deviation for priority grantees. The specification of the model within a Bayesian format allows us to incorporate prior information from earlier studies of topic change for researchers. We generally know that researchers tend to focus on a narrow set of topics, and even if topic switching has increased recently, researchers rarely venture beyond their broader research area (Aleta et al., 2019; Zeng et al., 2019). Moreover, Myers (2020) show that the economic incentive built into targeted grants has to be quite substantial to induce shifts in research topics. We therefore set somewhat sceptical priors on the group means. We specify two priors for the two group means in the form of either a truncated normal distribution $Normal(0.7, 0.2)$ with lower bounds at 0, and upper bounds at 1, or a beta prior $Beta(1, 0.6)$. Because we remain sceptical of a difference between groups, we set the same prior for both non-priority and priority grants, which translates into a prior on the group differences centred around 0. Recall that the similarity measures are 1 for no change in topic distributions, so this puts the bulk of probability mass around values that indicate little change. For all models we assume a normally distributed prior of $Normal(0, 1)$ for the variance parameter σ.

However, the bounding of the similarity measures between 0 and 1 may also provide a problem for recovering accurate group means. For all similarity measures, some portion of grant holders engage in exactly the same topics or completely different topics. The distribution of topic changes are thus inflated with 0's and 1's. We therefore supplement the analysis with a generalised linear model, using a zero-one-inflated beta regression (ZOIB). The model uses a mixed response distribution with a beta distribution for data within the closed (0, 1) interval and a Bernoulli distribution for the binary $0, 1$

responses. The response distribution is therefore parameterised with four parameters, such that:

$$\begin{aligned} S &\sim ZOIB(\alpha, \gamma, \mu, \phi) \\ logit(\alpha) &= \pi_0 + \pi_1 P_i + \epsilon \\ logit(\gamma) &= \tau_0 + \tau_1 P_i + \epsilon \\ logit(\mu) &= \beta_0 + \beta_1 P_i + \epsilon \\ log(\phi) &= \delta_0 + \delta_1 P_i + \epsilon \end{aligned}$$

Where α denotes the probability of a similarity of either 0 or 1, γ is the conditional probability of a similarity of 1 given that it its either 0 or 1, μ is the mean for values not 0 or 1, and ϕ the precision parameter. Because the group means are estimated on the logit scale, priors from the linear model would be overly restrictive. We estimate the ZOIB model with two different priors on the group means: $Normal(1, 1)$ and $Student - t(3, 1, 1)$. For the three new parameters, we use a $Beta(1, 1)$ prior for α and γ, and a $Gamma(0.01, 0.01)$ prior for ϕ. We estimate 72 models of topic similarity based on three samples (A, B, and C), three outcomes (before vs. during, during vs. after, during vs. after), two matching methods (CEM and IPW), two model specifications (linear and ZOIB) and two priors (Normal vs. Beta and Normal vs. Student-t).

Finally, we use a linear model for comparing the diversity between priority and no-priority grants. However, we use repeated measures of diversity for each grant, and therefore account for this through a hierarchical model with random intercepts for each grant

$$\begin{aligned} D &\sim Normal(\mu, \sigma) \\ \mu &= \beta_0 + \beta_{grant_i} + \beta_1 P_i + \epsilon \end{aligned}$$

with a $Half-Normal(0, 0.1)$ prior for each group mean (β_0 and β_0+β_1), and $student-T(3, 0, 10)$ for both σ and β_{grant_i}. All models were fitted through a Markov chain Monte Carlo (MCMC) simulation, using a Hamiltonian Monte Carlo algorithm implemented in the brms package (ver. 2.12.0) (Bürkner, 2017) in R (ver. 3.6.3). Four parallel chains were run with 1,000 iterations used for warm-up, 4,000 total iterations, and 3,000 iterations used for calculating posterior densities.

9.5 Supplementary material

9.5.1 Success rates for priority and non-priority grants

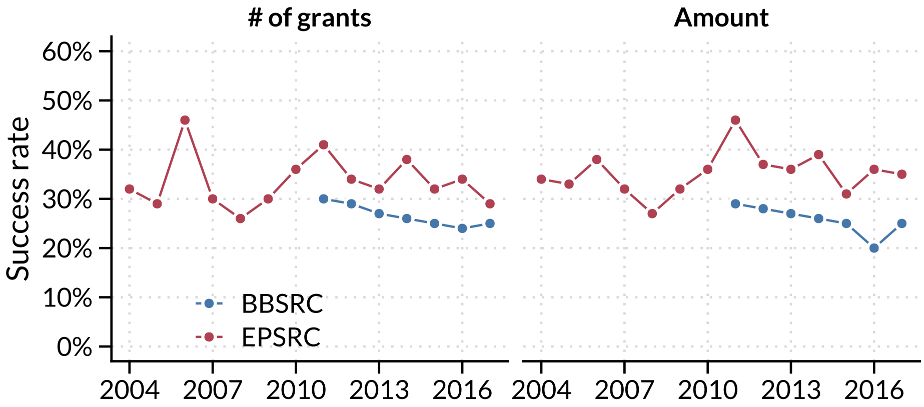


Figure S9.1: Success rates in BBSRC and EPSRC. Total grant success rates for number of grants and amount of funding.

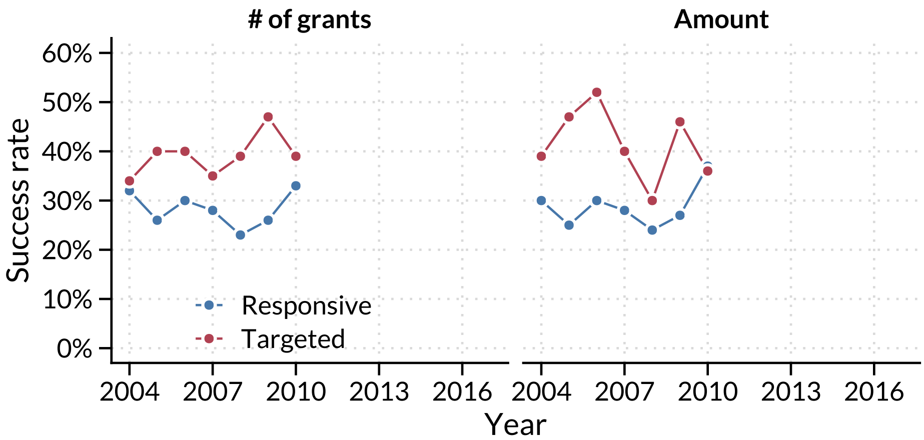


Figure S9.2: Success rates in BBSRC and EPSRC. Success rates for number of grants and amount of funding by grant mode.

9.5.2 Covariate balance for matched and unmatched groups

Table S9.1: Group balance for numerical and categorical variables. Sample A.

	Unmatched	Matched
Total publications	0.315	-0.232
Total normalized citations	0.793	-0.382
Median normalized journal impact factor	0.025	0.001
Academic age	-0.338	-0.046
Funding organization*	66.38	1.534
Type of grant*	6663	2472
Most prevalent discipline*	140.1	56.36
Multivariate imbalance measure (L1)	0.975	0.915

Note: Average group differences for numerical variables are treated – untreated. For categorical variables*, the difference refers to the Chi-square difference.

Table S9.2: Group balance for numerical and categorical variables. Sample B.

	Unmatched	Matched
Total publications	0.191	-0.208
Total normalized citations	0.774	-0.075
Median normalized journal impact factor	0.026	0.002
Academic age	-0.468	-0.043
Funding organization*	54.68	0.981
Type of grant*	5350	1818
Most prevalent discipline*	127.8	57.49
Multivariate imbalance measure (L1)	0.978	0.925

Note: Average group differences for numerical variables are treated – untreated. For categorical variables*, the difference refers to the Chi-square difference.

Table S9.3: Group balance for numerical and categorical variables. Sample C.

	Unmatched	Matched
Total publications	0.925	-0.349
Total normalized citations	2.097	-0.267
Median normalized journal impact factor	0.021	0.001
Academic age	-0.457	-0.030
Funding organization*	38.50	1.995
Type of grant*	3409	749.0
Most prevalent discipline*	108.3	50.29
Multivariate imbalance measure (L1)	0.977	0.908

Note: Average group differences for numerical variables are treated – untreated. For categorical variables*, the difference refers to the Chi-square difference.

9.5.3 Additional descriptive findings

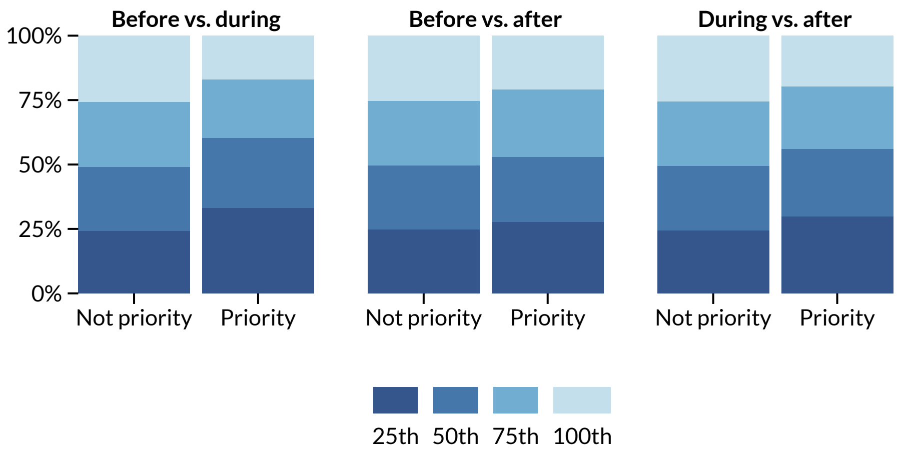


Figure S9.3: Topic change quantiles for priority and non-priority grants. Columns denote the percentage of grants within a group that fall within a certain quantile of topic change for each change comparison.

9.5.4 Robustness tests

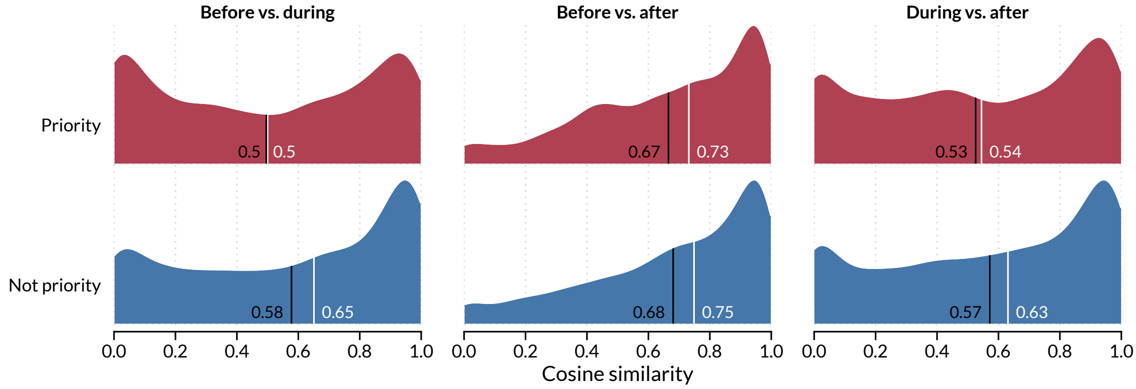


Figure S9.4: Topic-similarity distributions for priority and non-priority grants. Vertical black lines specify within-group mean similarities. Vertical white lines specify within-group medians. N = 10,475. Figure is identical to figure 9.1, but uses a more fine-grained topic delineation with 3,972 individual topics.

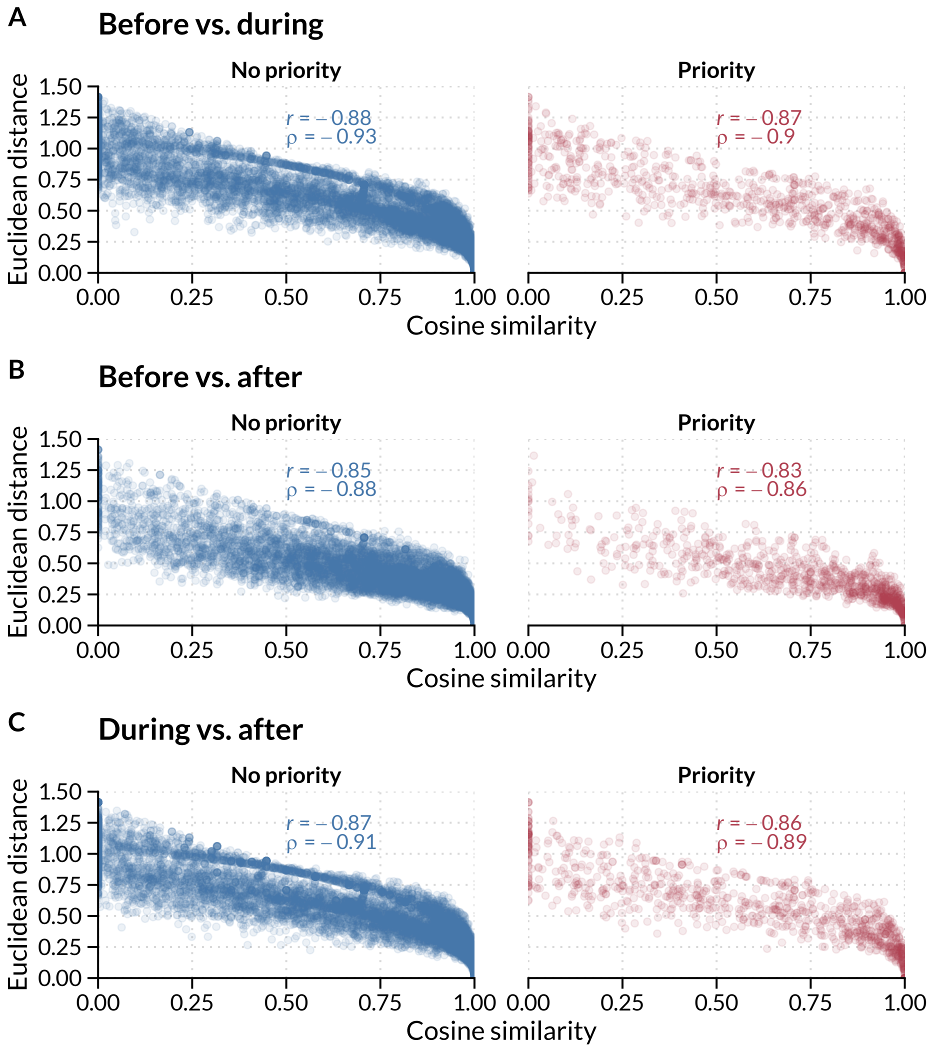


Figure S9.5: Similarity measure correlation. Correlations between cosine similarity measure and Euclidean distance in sample A with N = 10,475. Annotations show Pearson's correlation coefficient (r) and Spearman's rank correlation (ρ).

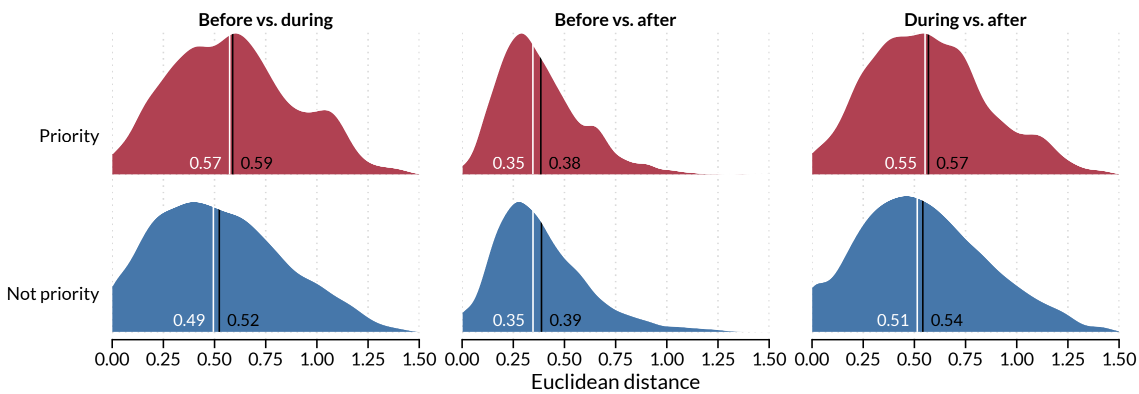


Figure S9.6: Topic-similarity distributions for priority and non-priority grants. Vertical black lines specify within-group mean similarities. Vertical white lines specify within-group medians. N = 10,475. The figure is identical to figure 9.1, but uses the Euclidean distance to measure topic similarity. Contrary to cosine similarity, higher values indicate less similarity.

10. The Interdependence of Research Funding Concentration, Policy Priorities, and Problem Choice

Emil Bargmann Madsen[1]

The prioritisation of research funding towards a small elite of researchers and research topics of "strategic" importance are becoming a norm across national research systems. Researchers are increasingly worried that such steering hampers the diversity of scientific approaches and problems addressed. However, the effects of increased steering of who and what receives research funds are not well known. I use evidence from 65,000 research grants awarded by 7 research councils in the United Kingdom and 15 Danish research funders to investigate how strong funding concentration and thematic targeting leads to less topical diversity. Researchers in the very top of the funding distribution primarily investigate topics and disciplines with the most funding success, and research output form targeted funding schemes overlaps with that from investigator-led grants. Moreover, priorities from private funders line up with the type of research funded by public research councils. The findings highlight how steering through funding decisions can multiply across the scientific system and exacerbate existing funding concentration.

[1]Danish Centre for Studies in Research and Research Policy, Department of Political Science, Aarhus University

10.1 Introduction

Over the course of the past decade, the institutional structure of research funding have changed profoundly. This change have been especially visible in two areas: First, more academic research have become tangent on researchers' ability to secure funds and resources from sponsors external to their research organisation, including both public, private, and non-profit research funders (Whitley et al., 2018, p. 110). Combined with an increasing degree of competition for these funds and low success rates, funding is increasingly concentrated in the hands of a small elite of principal investigators and institutions (Berg, 2012; Bol et al., 2018; Katz & Matter, 2020; Ma et al., 2015; Mongeon et al., 2016).

Second, the greater reliance on competitively- and externally evaluated grants have propelled funding decisions and priorities to the forefront of research policy tools for governments and other funders alike to utilise. This takes place in a direct way, as funders have sought to steer research in particular directions by allocating grants towards favoured research areas, problems, and approaches (Gläser & Laudel, 2016; Gläser & Velarde, 2018). Private funders, such as pharmaceutical or biotech companies, operate on a strict market-based logic, investing in research problems with a profitable potential (Jones & Wilsdon, 2018; Yegros-Yegros et al., 2020). Even if these firms have diversified their research portfolio (Ràfols et al., 2014, p. 27), biomedical research priorities remain tethered to the market for treatments rather than societal needs (Evans et al., 2014; Yao et al., 2015). Governments and public funding bodies have, on the other hand, attempted to ensure that such needs are prioritised in funded research by targeting areas and problems deemed "societally relevant". This has been the case in the UK, where research councils have adapted to new government priorities through greater activism in funding of selected areas (Kearnes & Wienroth, 2011; Nature, 2003), or in Denmark where new strategic research councils and technology-oriented foundations were tasked with selecting projects that could foster innovation and solutions to society's problems (Aagaard, 2017).

While this *direct* type of steering science towards certain research problems have garnered much interest, there is very little quantitative evidence of how funding concentration and policy-induced prioritisation in funding influences the direction research (Ciarli & Ràfols, 2019; Gläser & Laudel, 2016). How research content, including problem choice, is linked to changing funding practices is a pertinent question but still underexplored (Laudel & Gläser, 2014, p. 1204; Whitley et al., 2018). The mechanisms governing how funding conditions influence both individual researchers' problem choice and the research system's wider topic and disciplinary diversity are likely many and interwoven. Without detailed knowledge of how scientists perceive and adapt to scarcity of research funds and the steering hereof, we cannot meaningfully separate the direct effects of funding arrangements from other incentives and constraints inheret to the scientific system.

I instead focus on the ways the growing concentration of research funding and funder prioritisation have facilitated greater steering in *indirect* or even *unintended* ways. I focus on two mechanisms related to the increased prioritisation of research funding. I ask:

1. Does stark concentration of resources for individuals coincide with a concentration within research problem/disciplinary choice?

The tendency for funders to increasingly prioritise few, ”excellent”, researchers are well established, and concerns for how it impacts the type of research conducted are prevalent. Fewer, and larger, grants could narrow the diversity of projects, induce more conservative assessment of grant proposals, and make applicants focus on problems with more certain outcomes (Bloch & Sørensen, 2015, p. 36; Alberts et al., 2014; Fortin & Currie, 2013). Moreover:

2. How do prioritised, strategic, funding instruments in public research councils, and priorities from private funders impact ”free”/investigator-initiated research?

Both prioritised public and private funding have been speculated to spill over into traditional grant applications by incentivising an overemphasis on certain problem areas (Bloch & Sørensen, 2015, p. 36-37; Yegros-Yegros et al., 2020, p. 3).

To study this, I use information on approximately 65,000 research grants awarded by seven UK research councils, and 15 public and private Danish research foundations. I focus on the strong concentration of funding for individuals and topics at the system-wide level, but also investigate what individual researchers choose to study, and how it relates to the amount of funding they receive and from what instruments and funders they recieve it.

I document that funding is indeed concentrated in both few individuals, research disciplines, and topics. To contextualise this finding, I also show that the well-funded principal investigators mainly engage in well-funded topics, and grantees seldom diversify their disciplinary and topical focus from grant to grant. Irregardless of the causal direction, funding seems to be biased towards more prestigious topics. Furthermore, I show that both private and public research funders share priorities, and tend to fund the same topics the most. This prestige overlap could stem from the fact that the most successful grantees often receive funding from a more diverse set of funders.

These potential biases, and underlying perpetuating mechanisms, have major implications for research policy. It lends credence to already existing arguments in favour of dispersing research funding to obtain more diversity in research areas and approaches (see Aagaard et al., 2020, for a thorough discussion). Moreover, recent suggestions for implementing a modified lottery in grant competitions (Fang & Casadevall, 2016) could potentially remedy the strong dual concentration for individual researchers and research problems. The paper makes several contributions to the brewing discussion of how to sponsor research, by illuminating the ways policy decisions about increased concentration and prioritisation of research funding may inadvertently exacerbate existing inequalities in the science system.

In the following, I revisit the existing evidence on the intersection of problem choice and funding concentration. Next, I lay out a rationale for how funding concentration on individual researcher may impact the spread of studied research problems at the systemic and individual level. I then describe the methodological base for the analysis of Danish and UK grants, report on this, and finally discuss the implications of increased concentration for science policy.

10.2 Interdependence in funding priorities

The general scarcity of research funding budgets have long made it necessary for research funders to prioritise who they fund, and what type of research they support (Stewart, 1995; Waltman et al., 2019). Early accounts of how different lines of research are prioritised in the scientific system have generally framed it as a rational process where outcomes are selected in an optimal way given budget constraints (see Ciarli & Ràfols, 2019, p. 952, for a discussion). The assumption have been that the research community is self-organising and self-regulating, and researchers will tend to allocate attention to the most important topics when left to their own devices (Polanyi, 1962; Sarewitz & Pielke, 2007). In Polanyi's (1962) vision of science as a republic, scientists pursue problems and adjust their problem choice by taking into account what other researcher have worked or are working on. Prioritisation of different topics then emerges as guided by an invisible hand: Researchers self-coordinate, which leads to a joint result no individual intended. Accordingly, this self-coordinated focus on certain topics should ensure "the most efficient possible organization of science" (Polanyi, 1962, p. 54). As noted by Sarewitz and Pielke (2007), this is a very stylised view of how scientists choose what topics to work on, and few would argue that this extreme form of self-organisation takes place. However, many researchers and policy-makers may acknowledge that strategic investments in certain topics and disciplines are necessary, but societal needs are best addressed when researchers are relatively unrestricted in their pursuits (Sarewitz & Pielke, 2007, p. 7).

Contrary to this, I will argue that prioritisation of different topics and disciplines through funding cannot be understood as a relatively free 'marketplace of ideas', where funding is allocated to areas proportional to their importance in the research community. Empirical studies have shown that scholarly attention and also funding is highly stratified, with few areas amassing disproportionate resources compared to the rest. A wealth of studies in bio-medicine shows that research funding and publications are highly concentrated in few select areas. Grants from the US National Institutes of Health tend to focus on a narrow set of diseases with around 40 % of funding flowing to 10 % of diseases, and funding distributions in one year often mirror previous years (e.g. Gillum et al., 2011; C. Gross et al., 1999; Yao et al., 2015). This would not be unexpected from the perspective of science as a self-coordinating system that I presented above. However, many studies show that there is often a gap between societal needs and research priorities. There is high relative investment in disease areas most likely to affect the populations in high income countries (Yegros-Yegros et al., 2020), investments often match interest by pharmaceutical companies (Evans et al., 2014), and already well-studied genes make up the bulk of genes studied in NIH-funded projects (Stoeger et al., 2018, p. 7). At the same time, research funding is increasingly concentrated in the hands of few researchers. Across many funders, grant amounts have increased, but are simultaneously awarded to a small selection of individuals (see e.g. Katz & Matter, 2020; Ma et al., 2015; Madsen & Aagaard, 2020; Mongeon et al., 2016). Often, around 20-25 % of grantees recieve 75-80 % of funding, all the while leading to decreasing marginal returns in terms of publications and other research output (Aagaard et al., 2020; Mongeon et al., 2016).

I present two perspectives on how the concentration of funding in individuals and topics interact through different prioritisation mechanisms. First, prioritisation can be aimed at how science is conducted (e.g in large or small teams) or who is involved (a small 'elite' of researchers or a broader growth layer). This type of prioritisation takes place as funders increase grant sizes, while success rates decline and grant money become increasingly concentrated. Second, prioritisation can happen through select emphasis. Here, priority is placed on the substantive content of scientific research by designating particular disciplines or research topics, on which privileged status is bestowed on. Such thematic priorities are the most common way of thinking about research funding priorities (Hellström et al., 2017; Jones & Wilsdon, 2018; Wallace & Ràfols, 2018).

Below, I outline how these prioritisation mechanisms may impact the distribution of funding to different topics and disciplines in the aggregate, but also how it could affect the topic choice of individual researchers in a stratified manner. Where individual choices of research question affects the career trajectory of the researcher, scientists' combined choices affect the direction and efficiency of scientific discovery as a whole (Rzhetsky et al., 2015).

10.2.1 Individual funding concentration and topic choice

The connection between increasing concentration of funding and a lack of diversity in what topics and research areas are studied have received much attention and speculation. A high degree of systemic funding concentration is thought to inhibit risk taking in funding competitions. Funding a proposal can be thought of as an experiment with uncertain outcomes. When funding is concentrated on a small group it induces risk as the number of experiments are reduced (Bloch & Sørensen, 2015; Fortin & Currie, 2013). This can perhaps explain greater concentration of funding in few topics, as researchers choose topics with a proven track record, which in turn creates a more static and less diverse system (see Aagaard et al., 2020). This would explain why e.g. funded projects focus on already well-studied genes an molecular entities, and not those with more unknown functions (Edwards et al., 2011; Stoeger et al., 2018). However, without access to the inner thought process of researchers when applying to funding, we do not know whether this is the case.

Instead, we can think of the prioritisation of already well-funded researchers as bringing broader organisational issues, which inhibit the diversity of topics and research areas funded. Funding concentration disproportionally affects early-career researchers, who cannot compete with the publication record and resources of a well-funded senior elite (Aagaard et al., 2020; Peifer, 2017). Moreover, it contributes to the broader cumulative advantage of certain privileged individuals and research groups (Allison et al., 1982; S. Cole, 1970; Merton, 1968). Cumulative advantage in research funding signifies that individual researchers receive differential levels of funding depending on earlier funding, recognition, and type of work they have done. For example, scientists studying the least popular quintile of genes are around 12 percentage point less likely to win subsequent funding compared to scientist focusing on the most studied genes (Stoeger et al., 2018, fig. S10). So when grant sizes increases but success rates are low [2], some successful research group have much better access to resources, which can be used in subsequent funding competitions. This can result in a continued stratified support for a narrow set of specialised research areas and groups working within particular fields. The possible cumulative advantage of well-funded grantees leads to a set of expectations about the funding distribution at the aggregate level and the level of the individual researcher.

[2] Between 2011 and 2019, success rates for obtaining a grant have fluctuated between 14 and 19 % in the Independent Research Fund Denmark (which includes the five research councils). In UK research councils, the success rate in e.g. the Biotechnology and Biological Sciences Research Council saw a drop from 30 to 25 %.

Aggregate funding distribution. If the distribution of funding to individuals and topics is interdependent, we should expect the cumulative distribution of funding to individual investigators to be considerably skewed, and the aggregate distribution of funded topics to be just as unequal. However, it is important to note that these systemic characteristics would not in itself suggest an interdependence of funding inequalities. The distribution of funding to topics could be unequal because all grantees focus on a smaller set of topics, in line with what would be expected from the self-coordination perspective. The interdependence should also be visible at the level of individual researchers.

Topic distributions across individual researchers. In order for cumulative advantage to drive the unequal distribution of funding to topics, I expect researchers with different levels of funding resources to investigate very different topics, and researchers in the top of the funding distributions should also focus on the topics with the most funding. Conversely, if the two distributions are independent, a skewed distribution of funding to topics will stem from researchers across the funding distribution researching similar topics.

10.2.2 Thematic priorities, private funders, and spill-over of topic-choice

The strong competition for funding, and its consequences for topic diversity, is a concentration mechanism *internal* to the scientific system. Greater competition and notions of what constitutes excellent types of research and researchers are likely to become internalised in funders' assessment criteria. The use of these criteria then signals to researchers, journals, and reviewers what topics are deemed prestigious (Madsen & Aagaard, 2020, p. 17). However, a simultaneous development in the funding system have been the growing *external* steering of research through funding allocations to thematic priority areas and problems (Gläser & Laudel, 2016, p. 126; Whitley et al., 2018, p. 110). This targeting of research funding have been driven by both public research funders and private companies, foundations, and charities. Crucially, targeted funding signals to the scientific community what areas and topics are important and where to expect future possibilities for funding. Targeted priorities are especially argued to solidify emerging fields such as nanotechnology or synthetic biology (Gläser & Laudel, 2016, p. 127), but then continuing to skew research efforts through continued investment in few strategic areas (Nature, 2003). Differences in assessment criteria can further amplify this skew. In the BBSRC, *"Excellence is the overriding criterion in the assessment of research grants (...)"*, but *"(...) it is expected that competitive applications that address a responsive mode priority will have some advantage in competition."* (BBSRC, 2020). In the Engineering and Physical Sciences Research Council (EPSRC) researchers have experienced much starker declines in success rates for investigator-initiated programmes than targeted programmes (Nature, 2003).

For private funders, the use of targeted priorities tend to be more ”acceptable” for the research community. Many private funders operate within certain market demands and focus on research with a profitable potential (Sarewitz & Pielke, 2007; Yegros-Yegros et al., 2020). However, because private funders in especially health and biomedical research have expanded their research investments this may affect publicly funded research too. When 60 % of health research is rooted in the private sector, and many pharmaceutical companies spend more money on research than e.g. the Wellcome Trust (Røttingen et al., 2013), funding from industry will likely influence the topic selection of researchers even if their studies are based on public funding. Pharmaceutical research priorities then “spill-over” into public research priorities (Yegros-Yegros et al., 2020, p. 3; Evans et al., 2014), perhaps because some areas are easier to attain funding for. Targeting by public and private funders can also lead to a set of expectations about both aggregate funding distributions, and individual researchers’ topic choices.

Aggregate funding overlap. Spill-over between targeting and private priorities to more ”free” funding programmes and instruments will entail a positive relationship in the relative distribution of funding to different topics and disciplines between the programmes. Put plainly, those topics and disciplines in the top of the funding distribution in targeted or private funding programmes, should be similarly placed in funding programmes or organisations with less topical binding.

Targeted grants and topic choice. It is important to note that an overlap in themes funded by public vs. private funders or targted vs. non-targted funding programmes may suggest spill-over, but not how these spill-overs manifest for individual researchers. Topic-choice is a complex task, and researchers likely select problems that strike a balance between their own interest, what the scientific community deem important, and what is expected by funders (Yegros-Yegros et al., 2020, p. 2). The impact of targeted funding on individual topic choice could play out in different ways. First, targeted funding opportunities could lead researchers in some areas to abandon or change their focus in order to match prioritised areas. Second, targeted programmes could instead privilege scientist already working in prioritised areas, giving them a competitive edge in other funding competitions. I return to all the outlined expectations in the analyses below.

Table 10.1: Summary of grant samples

Sample	Years	No. of grants	No. of PIs	Grant value[a]		
				Mean	Median	Std. dev.
Denmark	2005-2016	18,404	8,624	2,555,962	984,191.0	6,690,068
United Kingdom	2005-2017	41,416	22,290	434,732	239,807.2	1,425,052

Note: [a] Values are given in DKK and CPI-adjusted pounds respectively.

10.3 Data and methods

10.3.1 Data on research grants

To identify how funders allocate resources to different research topics and areas, I rely on detailed data on both funded research grants and the output produced from these grants in the form of scholarly publications. The data on grants come from seven public research councils in the UK, and 15 major research funders (incl. public and private) from Denmark.[3] For the UK sample, I scraped all available information on funded research projects from the AHRC, BBSRC, EPSRC, ESRC, MRC, NERC, and STFC in the RCUK Gateway to Research Database from 2005 to 2017, in the BBSRCs advanced grant search, and the EPSRC Grants on the Web database. For the Danish sample, individual funders were contacted and asked to provide project-level data from 2004 to 2016. Table 10.1 provides an overview of the two samples of grants.

In both cases the number of grants do not provide a full picture of all grants awarded by the UK research councils or Danish research funders in that period. For both samples, many projects are not strictly research-related or aimed at producing a traditional scientific output. I have excluded all grants aimed at training or education, and kept pure research grants and fellowships. Furthermore, all grants were linked to the primary principal investigator only, and grants with a running amount paid by year were centered on the first year the grant was received. For the UK sample, I rely on the the individual investigator ID provided by the research councils to disambiguate individual PIs. Grants where a PI could not be disambiguated where kept out of the analysis, yielding 41,416/53,095 $\approx$ 78 % of the total number of grants. For the Danish sample, grant recipients were manually disambiguated and linked to grants. For common name combinations, the grant recipient was identified through a combination of data on their institutional affiliation and primary research area. If not successfully disambiguated through

[3]The funders are: The Carlsberg Foundation, the Independent Research Fund Denmark (DFF), the Danish National Research Foundation (DNRF), the Council for Technology and Innovation, the Strategic Research Council, the Advanced Technology Foundation, the European Research Council (ERC), the Danish Cancer Society, the Lundbeck Foundation, the Nordea Foundation, the Novo Nordisk Foundation, TrygFonden, the Ministry for Higher Education and Research, the Velux Foundation, and the Villum Foundation

this, grantees were considered separate individuals. Consequently, the sample has less comprehensive coverage of strategic and innovation-related funding, as these types of grants are often awarded to a consortium of firms and research organisations with no single discernible PI. These limitations are important in qualifying the results presented below. Despite this, the samples cover larges parts of competitive funding awarded over a twelve year period, and while they may not accurately depict the absolute funding amounts, they provide important insights about the relative distribution of scarce resources in the research systems.

10.3.2 Categorising grants into topics and disciplines

For both the UK Research Councils and the Danish research foundations, we lack a consistent system for classifying grants into research disciplines and research topics. First, while the UK grants are often classified from a standardised scheme by the grantee themselves, the classifications schemes differ across councils. Second, these classifications seem to work on very different levels of aggregation with the MRC using 26 areas for classification while e.g. the EPSRC uses over 200. Third, we cannot compare across the UK and Danish samples without establishing a common classification system.

To consistently classify grants, I use journal articles and reviews by the principal investigators of each funded project. This do entail some methodological assumptions, as I expect a principal investigator's publications to reveal the underlying mix of topics and disciplines of the grant. I also assume that the topic mix has not changed much from grant proposal to output. Using the publication output also means that I am assessing the type of research that a grant actually resulted in, instead of just its intended results. For grants in the UK sample, the Gateway to Research database provides DOIs of publications reported by the grantees themselves as being outcomes from their grants. I match these DOIs to records in the CWTS in-house version of Web of Science (WoS). For Danish grants, I combined automated string and manual matching, by comparing names, email addresses and institutional affiliations in WoS with the grant data. Here, I cannot discern what publications were actual outputs from a grant. Instead, I consider the publications 0–4 years after a grant was received. None of the samples were completely covered by the WoS data, but the matched grants account for around 70% of the combined funding amounts attributed to individual principal investigators. Table S10.1 in the supplementary material shows comparisons between the full samples and the WoS-matched samples.

Using the matched publications, I categorise grants into two different levels of aggregation, which merit further explanation. First, I situate each grant within a scientific topic. A topic can be thought of as the lowest level of aggregation in a hierarchical, content-focused, categorisation scheme. I understand a research topic as a cognitive and semantic space populated by a collection of documents focusing on a common problem area, and with a set of researchers engaged in this problem area. A topic is cognitive because researchers identify as parts of this topic and group text based on this, and semantic because documents and researchers share a common vocabulary and understanding of key terms within this topic. To identify the research topic of a grant, I use , I use algorithmically delinated topic clusters. The clusters are created by the Centre for Science and Technology Studies (CWTS) at Leiden University, and used in their Leiden Ranking of universities.[4] The topic of a publication is based on the direct citation links between the publication and all other publications in the WoS. Each article constitutes a node in a network with direct citations between papers constituting an edge between a pair of nodes. A Smart Local Moving (SLM) algorithm (Waltman & Van Eck, 2013) is applied for community detection and clustering into topics. Individual article nodes are moved into a partition and then aggregated so that each topic constitutes a node in an aggregated network. This procedure is outlined in (Waltman & van Eck, 2012) and (Traag et al., 2019). In all categorisation of research topics, a crucial question is the choice of granularity or resolution, i.e to choose how many different topics to include. I use two levels of granularity: a set of 3999 micro level topics, and a set of 864 meso level topics. These topics are hierarchically nested such that each micro topic is assigned to one meso topic, and each meso topic comprises a set of related micro topics. Second, I also place each grant in a set of 250 disciplines or specialities, based on the WoS Subject Classification. Disciplines are more aggregate categories of the scientific structure than topics, and denote a self-organised network of scientists working within the same topics, publishing in the same journals and readings/citing each others work (Sjögårde & Ahlgren, 2020, p. 210). I use the original classification of papers into disciplines, where each journal is classified, and a paper receives the same classification as the journal it was published in. However, this poses a problem because not all papers within a journal necessarily address the same discipline. To mitigate this, I follow Milojević (2020), and categorise a paper based on the most frequent discipline present in a papers cited references. This also eliminate a categorisation of grants into the residual ”Multidisciplinary Sciences”-category, as papers are assigned the most prevalent category from their references expect for this.

These four types of topic and discipline categorisations are by no means ”the correct” classification, but there are likely no ”ground truths” about how to classify a set of documents. The four categorisations I use instead fit the analysis in this paper by providing both different levels of aggregation and highlighting consistencies and inconsistencies across them.

[4] https://www.leidenranking.com/information/fields

The final categorisation of grants, places each grant into a mix of topics (two levels of aggregation) and disciplines (two methods), by weighting a grant according to its mix of publications. For each grant, I weigh both the grant sum and the grant itself according to the proportion of publications within each discipline or research topic. A grant of £100,000 with two publications within cell biology (66.6%) and two publications within virology (33.3%) is divided so that 0.66 grants and £66,666 are attributed to the former, while 0.33 grants and £33,333 are attributed to the latter.

10.3.3 Case selection and data considerations

I use data from both Danish research funders and UK research councils because each grant sample allows me to investigate different research questions. The data on Danish research grants cover a large section of the research funding system, and includes grants from private research foundations with ties to different industries (e.g. pharmaceutical companies, a brewery, and a window manufacturing company). This broad coverage makes it possible to compare how funders from both the public and private sector prioritise different research topics and disciplines, or the degree of overlap in their priorities. Furthermore, the inclusion of private funders, regular research councils, and strategic research councils provide a more encompassing view of the funding system and the distribution of resources across it. The UK data does not support such an analysis, but carries other advantages. The data from the UK research councils can be directly linked to research publications produced within the scope of each individual grant. This makes it possible to separate which topics and disciplines each grant engages in, and compare several grants awarded to one PI. The cases thus provide different opportunities, and I do not explicitly compare the cases in order to highlight any systematic and theoretically meaningful difference. Instead, the results show many commonalities between the cases, which may indicate some robustness of the findings.

10.4 Results

While much speculation have linked the increasing degree of funding concentration on a small group of "excellent" or "elite" researchers, and the tendency for funding to flow towards a narrow selection of research areas, disciplines, and topics, this interdependence is not well understood. Until recently, we knew very little about how research funders prioritise grant money, and how their selection criteria come to impact the type of research being pursued. Recent case studies in medicine (Yegros-Yegros et al., 2020), but also large-scale mappings of all areas (Waltman et al., 2019) show that funding is highly skewed towards certain disease areas and research topics. As outlined above, the missing piece of the puzzle is the underlying mechanisms driving these imbalances or

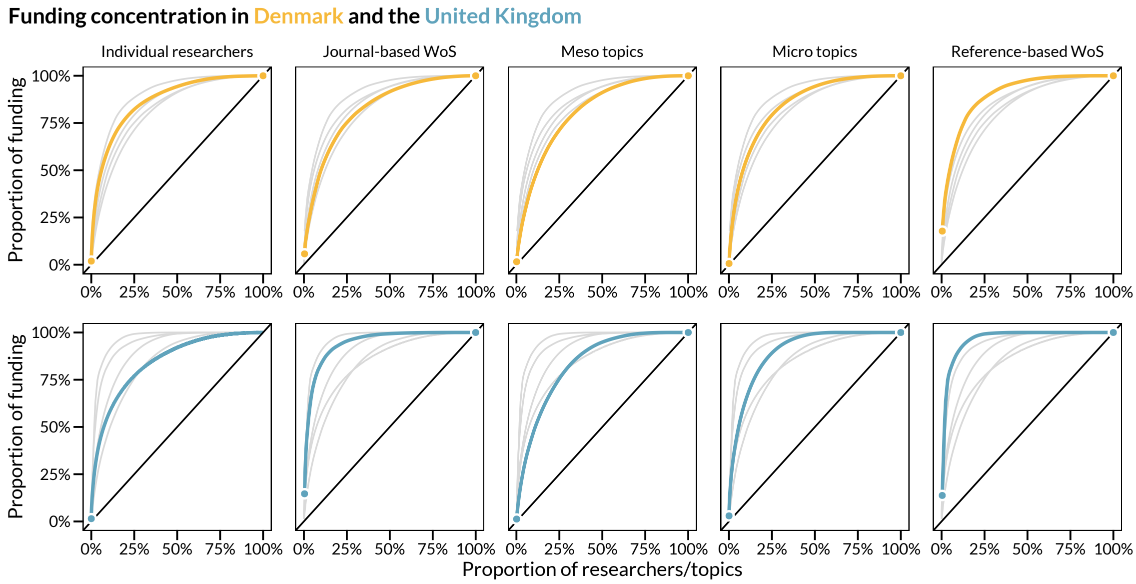


Figure 10.1: Funding concentration at the indvidual, topical, and disciplinary level

concentration trends. Such mechanisms are likely many, and may combine in complex and unpredictable ways. To get an overview of the degree of funding concentration, Figure 10.1 shows the Lorenz curves for funding across both individual researchers and the four levels of research topics and disciplines.

Broadly speaking, there tends to be a strong concentration of funding across researchers, topics, and disciplines and in both Denmark and UK. For individual researchers, the amount of funding is distributed so that 25 % of Danish grantees are awarded 82.2 % of all funding while 25 % of the UK grantees are awarded 77 %. This tendency for concentration is similarly strong across both micro- and meso-level topics. Across the 15 Danish research funders, 25 % of micro topics amass around 79 % of the total funding amounts in the years 2005-2016, while the concentration is somewhat reduced for the more aggregate meso-level with 71 %. For the UK, the degree of concentration is even stronger with a 25/76 and 25/89-division respectively. One explanation for the slight differences between Denmark and the UK could be the difference in funder composition. The Danish data are more encompassing and include grants from private research foundations. These grants may be directed at slightly different topics than what is funded by public research councils, as many of the Danish private foundations have more narrow scopes for their allocation defined in their charters. If we take a more aggregate look at the types of disciplines prioritised in the funding systems, we also find evidence for a very strong concentration trend. Using the regular WoS categorisation scheme, Danish funding is concentrated in a small handful of disciplines with 75 % of funding being directed at 25 %, or 63 disciplines, while 63 disciplines amass 95 % of funding in the UK. However, a lot of funding is directed towards "Multidisciplinary Sciences" because the original WoS categorisation bundles all multidisciplinary journals into one category. If we re-categorise these according to the most prevalent, non-multidisciplinary, category in the reference list we see an even stronger degree of concentration. For Denmark, the division is then 25/88 while the UK see an extreme 25/99 split.

Taken together, the distribution of funding amounts across individuals, topics, and disciplines show a strong stratification with few individuals and areas being prioritised, while the bulk remain scarcely funded. The simultaneous existence of skewed funding distributions indicate that at least some of the intuition from above could be true. Strong individual concentration could lead to topic concentration, or vice versa. It could also be that strong topic concentration is not due to few researchers getting a lot of funding for specific topics, but that all funded researchers engage in popular topics as a consequence of e.g. strong specialization, choosing topics based on opportunity, or survivor ships bias where only researchers engaged in popular topics stay in academia, get funding, etc. To more closely examine the interdependence of these funding distributions, we need to observe how individuals in different strata of the funding distribution choose topics.

10.4.1 Topic and funding concentration at the researcher level

So while the system-wide concentration of funding at both the individual and topical level is strong, a lack of diversity in who is funded does not automatically translate into less diversity in what is funded. If a wide selection of researchers engage in the same disciplines and topics, strong funding stratification is merely a sign of many grants (both small and large) flowing towards these areas and a relatively narrow focus from all grantees. This would create a positive self-reinforcing feedback for well-funded topics, but not necessarily a negative self-reinforcing feedback for low-funded topics. The collective topic choice could shift when new topics become popular or prestigious. Conversely, when a narrow elite of researchers engage with certain topics in a select range of disciplines, and relatively poorer researchers engage in different topics, with less funding success, the interdependence creates negative spirals from some researchers and topics.

Figure 10.2 ranks each grant recipient into 20 percentile ranks according to the cumulative amount of funding awarded to that person. This gives us 20 equally sized groups of researchers from different parts of the funding distribution. Next, we can calculate the proportion of researchers in each percentile rank group (e.g. the top 5 % funded researchers) who also work with a top 5 % funded topic or discipline, and compare it to the proportion of researchers across all funding ranks working within a top 5 % funded topic.

Across the Danish and UK case, we see a common pattern: More well-funded researchers tend to work within well-funded topics and disciplines, while researchers further down the funding distribution have lower propensity to engage in these. In the Dan-

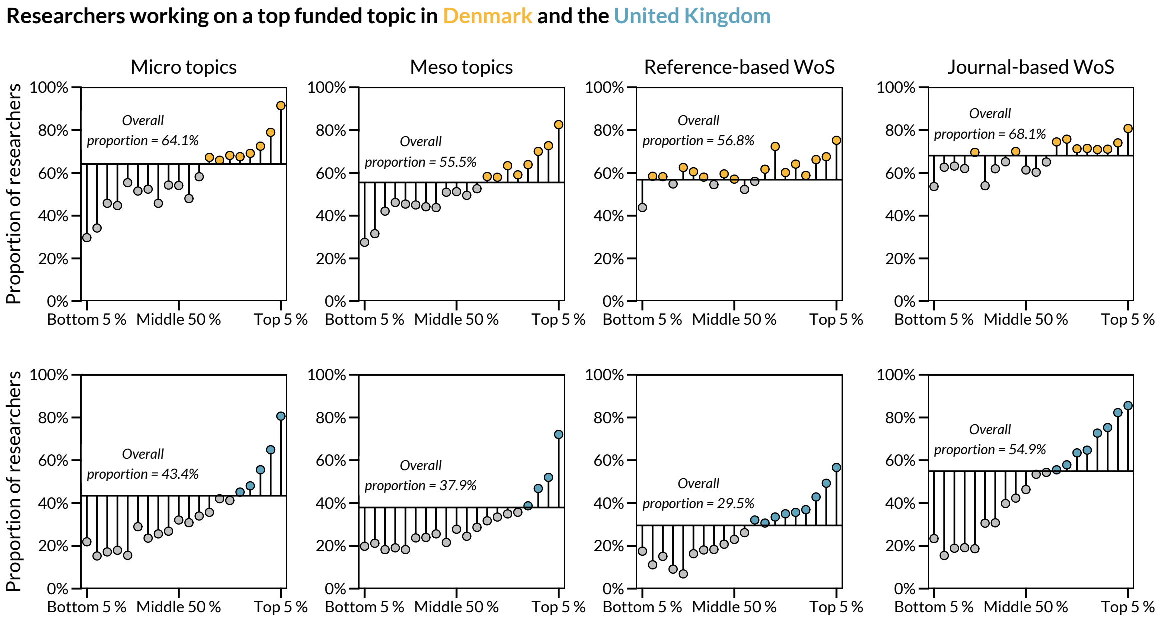


Figure 10.2: Individual funding concentration and propensity to work in top 5 % funded research fields

ish case, researchers across the funding distribution tend to work in top funded topics and disciplines, with 55-68 % of grantees addressing a top 5 % funded topic or discipline in some way. Looking at the disciplinary level, some differences across funding ranks exist but there seems to be very little structure to these differences. Generally, 60-80 % of researchers within the top 50 % of the funding distribution work in the most well-off disciplines, where this is just the case for around 40-60 % in the bottom part. However, large differences emerge at the topical levels. For meso-level topics, over 80 % of the top funded grantees work on the most funded topics, while less than 30 % of the least funded grantees work with these. At the most fine-grained topic level this difference is even bigger (insert difference). There are many possible interpretations of these patterns. First, I seems that funded researchers in Denmark are fairly specialised with many grantees working within a narrow set of topics and disciplines. Second, the lack of stratification in disciplinary focus vs. the strong stratification in topic focus is interesting. One explanation can be that biomedicine and clinical medicine are particular areas of funding and specialisation in Denmark (see Madsen & Aagaard, 2020, p. 12), but even within these areas a select group of problems and diseases are heavily funded. For example, diabetes mellitus drew approximately 11.8 % of disease-realted research funding in the period under consideration (Madsen & Aagaard, 2020, p. 13).

In the UK, the pattern of stratification is quite similar in all conceptualisations. For micro-level topics, the difference between the top 5 % and bottom 5% is almost 59 percentage points. In the top, over 80 % of grantees work on top funded topics whereas only 22 % of the bottom part of the funding distribution do the same. For meso-level topics, the difference is 52 percentage points, 50 percentage points for reference-based disciplines, and 62 percentage points for journal-based disciplines. In contrast to Denmark, the disciplinary and topical focus of UK grantees are less concentrated on heavily funded areas. The overall proportion of researchers working in these fields are around 30-55 %, but also varies much more between the disciplinary and topical level. It is interesting to note the difference between the reference-based categories and the original WoS categorisation. For the former, only 30 % of researchers focus on well-funded topics but the latter categorisation changes this to 55 %. A likely explanation is the heavy focus on prestigious, multidisciplinary, journals in many disciplines. When using the reclassification proposed by Milojević (2020), a grant's papers in Nature, Science, PNAS, etc. are redistributed to other disciplines.

Another contributing factor to the large degree of topical stratification can be that researchers tend to work on very related topics throughout their career. When the well-funded work on the most prioritised topics, and continue to do this, the funding concentration reproduces across time. Figure 10.3 compares the topical and disciplinary focus of a PIs first grant to all subsequent grants (for researchers with > 2 grants). This is possible with the UK data, as we can specifically link each grant to their publication outcomes, and thus distinguish the topical focus of individual grants. The measure for similarity is computed as the cosine similarity between a vector of topics addressed in a PIs first recorded grant (A), and a vector of topics from grants acquired 1, 2, 3, ..., n years after this (B). Each cell of these vectors take on a value equal to the proportion of papers from a grant that addresses a certain topic, such that topics not addressed equals 0, while topics adressed in a quarter of the publications equals 0.25:

$$\text{similarity} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{10.1}$$

where A_i and B_i are a topical or disciplinary proportion.

In the plot, a similarity score of 1 indicates a perfect overlap in topic or discipline profile between two grants, while 0 denote no overlap. Because the distribution of similarity scores are skewed towards 1, I show both the mean and median score. In general, a a PIs grant tends to be very similar in scope with mean and median similarity scores between 0.5 and 0.75. Naturally, a researcher's grants tend to be more similar at the disciplinary level than the topical level, and the similarity decreases with the number of topics we use to categorise. So while early grants seem very predictive of later topical focus, researcher do have room to shift topical focus within a broader set research agenda. Moreover we see a slight tendency for topic and disciplinary similarities to lessen as time goes by, indicating that researcher are branching out later in their careers. This largely corroborates findings from other analyses of topic switching behaviour (Horlings

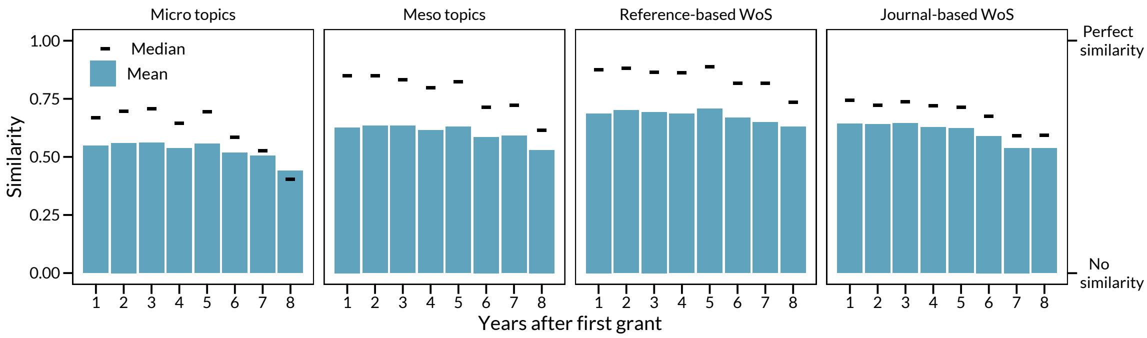


Figure 10.3: Topical similarity of grants across a PIs career in the United Kingdom.

& Gurney, 2013; Zeng et al., 2019). However, the similarity scores are only based on a sub-sample of researchers and may not be representative of the wider set of grantees.

In sum, the distribution of funding to topics, disciplines, and individual researchers are clearly linked with strong indications that highly funded researchers also research the most prioritised or popular topics. This stratification in topic focus across the funding distribution is likely one mechanism driving the huge inequalities in what type of research is funded in both Denmark and the United Kingdom. Furthermore, the strong path dependency of early projects could possibly contribute to a persistent strong concentration, as researchers across the funding distribution rarely change topic or disciplinary focus.

10.4.2 Overlapping priorities in diverse funding landscapes

The preceding analyses show how a prioritisation of funding towards a small segment of the researcher, or grantee, population may serve to increase the concentration of funding towards a handful of topics and disciplines. Another commonplace explanation of the interdependence of funding concentration for individuals and topics, are the specific topical priorities set by research funders. Recall that thematic priorities come in various implementations. First, public research funders have generally embraced the thought of earmarking funding to specific disciplines or topics deemed societally relevant, and have set aside funding instruments to encourage more research in these areas. Second, private funders have long focused on areas that are advantageous to their stated goals in charters or market strategies. Both types may lead to a spill-over of research priorities by incentivising research in specific research areas, and making it possible for already successful researchers to pull funding from many different sources.

Priorities in public funding. In figure 10.4, I aim to compare the relative priority or emphasis that strategic research councils or prioritised funding instruments put on certain topics, to relative priority given by traditional councils and instruments. Instead

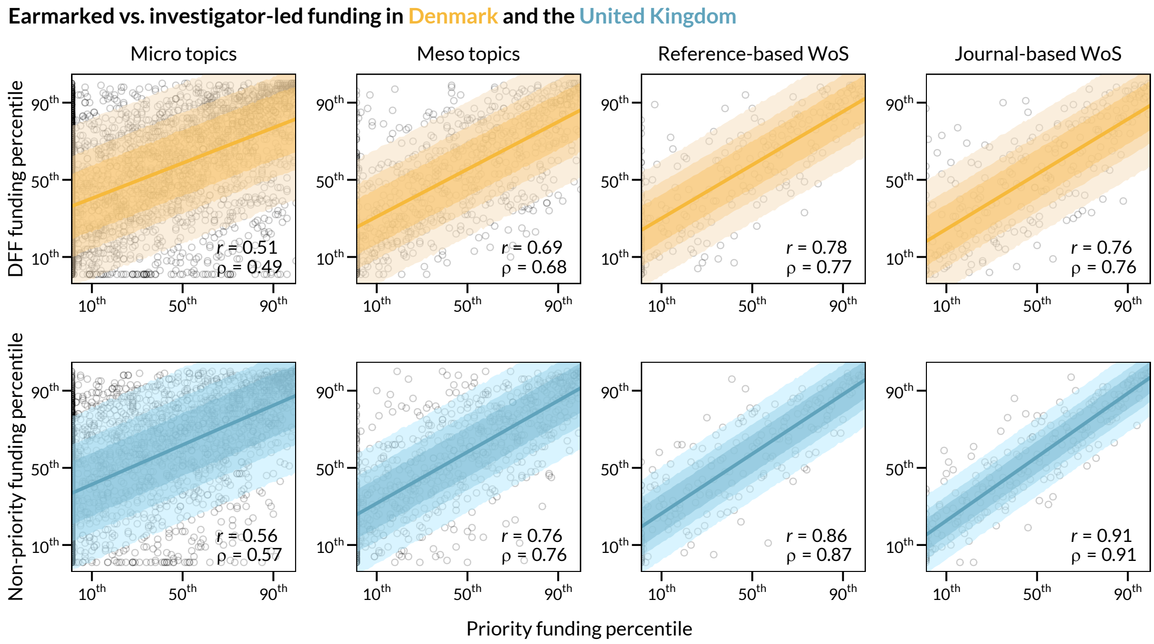


Figure 10.4: Topical overlap of funded strategic and "free" research in Denmark and the UK. *Note:* Solid lines are regression fits from a Bayesian linear model, with 50 %, 70 % and 90 % credible intervals calculated from the posterior predictive distribution. r = Pearson's correlation and ρ = Spearman's rank correlation.

of comparing the sum of funding for each topic from different sources, I calculate each topic's funding rank for each source and investigate whether the same topic is highly prioritised in each domain. The figure shows the basic scatter plot of funding percentiles. On top of the raw data points, a Bayesian linear regression indicates the degree of correlation between funding percentiles in strategic vs. "free" research councils (Denmark), and priorities vs. non-priorities grants in the EPSRC and BBSRC (UK). The shaded areas are 50 %, 70 % and 90 % credible intervals calculated from the posterior predictive distribution.[5]

At the systemic level, the correlation between funding priorities in strategic vs. non-strategic funding councils and funding instruments informs us about the overlap in what topics and disciplines are valued (or not). This is of course an abstraction, which factors away the variation in how much money is invested in different types of research by different funders, and focuses on the overlap in priorities for the average topic or discipline. However, the models show that at the disciplinary level, priorities line up to a surprising degree. In both Denmark and the UK, the percentile rank of a discipline within targeted and non-targeted funding instruments correlate positively, with correlation coefficients between 0.76 and 0.91. The strongest bivariate relationship is between prioritised and non-prioritised grants in the UK, where a one percentile increase in funding from targeted sources leads to a 0.83 increase in percentile rank among non-prioritised sources

[5] See Section B in the supplementary material for a full outline of the modelling procedure.

(based on the original WoS categorisation). In the lower end of the spectrum this increase is only 0.69 in the Danish case (reference-based disciplines). With the model, we can use a discipline's funding rank within prioritised instruments to guess its average rank in non-prioritised grant competitions. The simplified model predicts that 50 % of disciplines in the 25th prioritised percentile will fall somewhere between the 28th and 53rd percentile for reference-based disciplines in Denmark, and 50 % from the 75th percentile will have a non-prioritised rank of 63 to 87. We should note that while the relationship seems fairly strong at the average, the prediction intervals incorporate all residual variance (not predicted) in a discipline's rank, and highlights that it is likely determined by many other unmeasured factors.

At the topical level, the average relationship is weaker and much more uncertain. For meso topics, a 1 percentile rank increase from targeted funding translates into a 0.61 (Denmark) or 0.66 percentile (UK) rank change in non-targeted funding, but here the predicted funding ranks are very uncertain. For the Danish case, a topic in the middle of the targeted funding distribution would have a 50% probability of attaining a non-targeted rank of 41 to 69. 70 % of predictions for an average topic at the middle places it in a percentile rank between 34 and 77. In the UK, the predicted rank interval is almost identical (70 % fall between 39 and 78), and for micro topics the band is even wider. So although there seems to be some overlap at the topical level, it is much more variant than for disciplines. One reason for the large uncertainty for many of the predicted percentile ranks comes from large differences between broad research areas. Section C in the supplementary material reports results from a model similar to figure 10.4, but with a hierarchical structure estimating separate predictions for five main fields (Biomedical and health sciences, Life and earth sciences, Mathematics and computer science, Physical sciences and engineering, and Social sciences and humanities). For e.g. reference-based disciplines in Denmark, a percentile rank change in targeted funding leads to a average change of 0.76 percentiles for the biomedical fields, but only 0.67 in physics and engineering. These differences are even more pronounced for micro topics, where the corresponding model coefficients are 0.59 in biomedical fields, 0.42 in physics, and only 0.25 in mathematics and computer science. These large field differences stem from the fact that many targeted programmes are implemented within certain parts of the natural and medical sciences.

The question is how the overlap in priorities impact researchers' problem choice. As shown in figure 10.3, researchers tend to change their topical focus rather infrequently. In figure 10.5, I show the median similarity between a researchers grant based on targeted funding (represented at year 0) and grants obtained before and after this grant in the UK sample. The figure shows a lot of variation in similarity both before and after receiving a grant with a thematic priority. Similarities are consistently high at the discip-

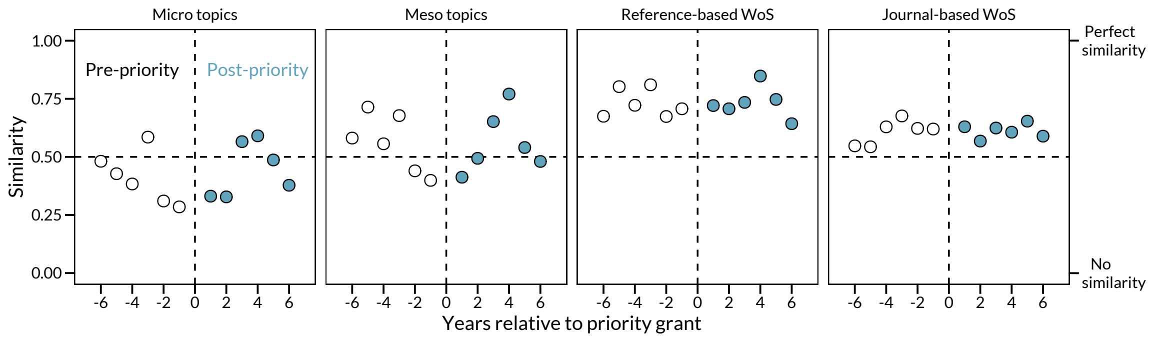


Figure 10.5: Topical similarity of grants for PIs with targeted funding.
Note: Points are median similarity scores relative to a priority grant at year 0.

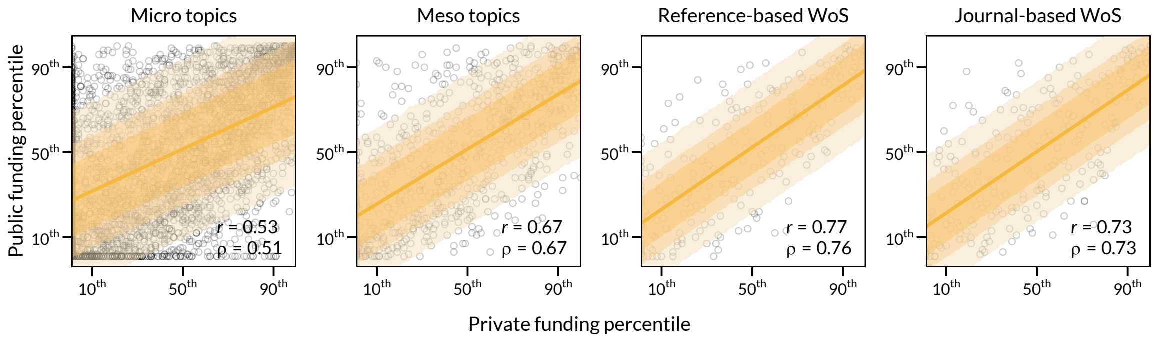


Figure 10.6: Topical overlap of funded private vs. public research in Denmark.
Note: Solid lines are regression fits from a Bayesian linear model, with 50 %, 70 % and 90 % credible intervals calculated from the posterior predictive distribution. r = Pearson's correlation and ρ = Spearman's rank correlation.

linary level, with scores between 0.7 and 0.8, while topical similarities tend to fluctuate more. For meso-level topics, the similarity is lowest one year before (0.399) and largest four years after (0.771). Interestingly, both topical and disciplinary similarities to a targeted grant tend to be highest 3-4 years before and after. This coincides with a median grant duration of 2.9 years, and may indicate that grants building on prioritised grant lead to a certain topic lock-in, but PIs are probably running multiple research lines in parallel.

Priority overlap between public and private funders. In addition to thematic priorities from public funders, there is a growing concern that the increase in funding from private funders contribute to a narrowing of funded topics and disciplines (Whitley et al., 2018; Yegros-Yegros et al., 2020). Below, I use information on grants from all the major public and private research foundations in Denmark to asses the overlap in funded topics. Figure 10.6 reports bivariate evidence similar to 10.4, but swaps funding percentiles from targeted sources to percentiles from private sources.

The overall pattern is similar to the one observed above. The correlation between funding percentiles are strongest for disciplines and lowest for the most detailed top-

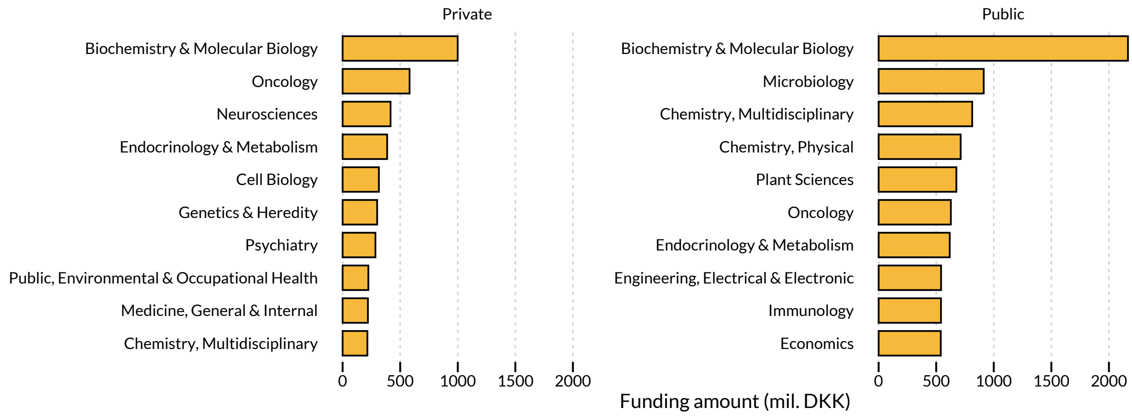


Figure 10.7: Amount of funding to reference-based disciplinary categories in Denmark.

ics. As evident by spread of individual topics and disciplines, the prediction of public funding based on private funding is also more uncertain as the categorisation of grants become more granular. For the reference-based disciplines, a one percentile score increase from private sources coincide with a 0.73 percentile increase in public funding. However, the prediction intervals reflect large individual variation for these disciplines. At the 25th private funding percentile, 50 % of the predicted public funding percentile lies within the [22nd - 47th] interval. Likewise, for disciplines within the 75th percentile of private funding, the model predicts a public funding percentile of 58 to 83. At the topical level, the regression coefficient for meso topics is 0.64 (95 % credible interval = 0.59-0.70) and 0.5 (95 % credible interval = 0.47-0.52) for micro topics. However, due to the greater variability for individual topics, the predicted public funding percentiles are more uncertain. For meso topics, % of predicted percentiles of a topic within the 50th private percentile are between 37 and 67, and for micro topics the range is 34 to 69. In the case of the Danish private funders, this lack of certainty stems from the funders strong focus on biomedical research. A number of the private research foundations are tied to pharmaceutical companies and invest the bulk of their funding budget in medicine, microbiology, and related fields. For the overlap between private and public funding, I also estimated a model which takes into account differences between the five major research fields (see section C in the supplementary material). In the domain of micro topics, the 1 private percentile increase is associated with an average 0.75 public percentile increase for the biomedical field compared to only 0.31 for the social sciences and humanities. Looking at the absolute amount of funding for different disciplines confirm the that private and public funders share at least some of their strongly prioritised areas. Figure 10.7 lists the top 10 funded disciplinary categories, and show that areas such as Biochemistry & Molecular biology, Cell biology, Micro biology and Endocrinology & Metabolism are high priority areas for both private and public funders.

10.5 Discussion

The results presented here provide an early empirical exploration of the role of funder priorities in skewing research towards a narrow set of scientific disciplines and topics. Data from both Danish and UK research funders confirm that competitively awarded grants tend to be accrued by few successful researchers, while the majority of grantees accumulate one to two grants over a prolonged period. Consistent with other studies of research funding (Madsen & Aagaard, 2020; Stoeger et al., 2018), the same degree of concentration exists for the type of research that is conducted within these grants. Funding is extremely skewed at the disciplinary level, but also for more fine grained topics or problems. The question is whether this dual concentration is somehow detrimental to the scientific ecosystem. Striking the right balance between concentration and dispersal of funding is a question of degree, not either or. Researchers generally agree that some measure of concentration is desirable (Aagaard et al., 2020). Not all research is of equal quality, not all problems merit the same level of attention, and some disciplines are naturally more resource dependent. A greater diversity of people and problems cannot become the end goal in itself as it may dilute all concerns for research quality and negate the need for prioritising scarce resources. However, the analysis above raises a number strong empirical indications of a system that has tipped too much in favor of concentration.

First, the analysis shows that the strong concentration on few individuals tend to contribute to less topical diversity because well-funded PIs selectively engage in the most prestigious and highly funded topics. A highly stratified distribution of funding is not in itself a problem, if funding is concentrated on a broader set of disciplines and result in research across a broad spectrum of problems. However, when funding organisations focus on individual applications, and delegate quality assessments to a peer review system, we run the risk of having multiple independent funding decisions skew funding in the aggregate. One problem might be that basing funding allocation on strict quality criteria, including past research performance, neglects a broader perspective on the whole funding portfolio. Individual review panels have little overview of what is funded in other panels or application rounds, meaning that high performing researchers continue to prevail. Different strategies for mitigating this lack of overview have been proposed. Directors at the National Institutes of Health have proposed a funding cap for individual investigators, of no more than three R01 grants at one time. This policy would supposedly free up funds for 1600 new grants. However, the policy were viewed with strong skepticism by many highly productive PIs and labs, and was subsequently abandoned. We could instead envision less invasive policies for highlighting and monitoring biases connected to a lack of oversight. In order to encourage topical diversity, more funding could be directed towards e.g. Discipline Hopping Awards or grant instruments aimed a building capacity in areas a PI have not explored earlier. This seems pertinent given that researchers tend not to change topics very often, and would likely need large funding incentives to do so (Myers, 2020). Grant proposals should also highlight how a project differs from earlier funded projects, review processes could be completely blinded, or funding could be allocated through a lottery of proposals deemed of sufficient quality (Fang & Casadevall, 2016). A broader portfolio of grant instruments with different aims and notions of quality will be necessary to broaden the type of research funded.

Second, priorities from targeted funding instruments appear similar to what is funded in through regular funding instruments. It is likely that funders have the prestige and perceived opportunity of different research topics and areas in mind when they set specific priorities and priorities are also set with input from the research community. As a result priorities may be suggested by successful and influential researchers, and/or reviewed and evaluated in councils and working groups where such researchers participate. This reflects a strong power dynamic where a select group of decision makers influence what type of research is deemed a priority. Moreover, because topic choice is hard to influence through priorities (Myers, 2020), the most effective instrument is rather funders' decisions not to fund certain topics (Madsen & Aagaard, 2020). When thematic and non-thematic funding instruments overlap this becomes a vicious cycle. If funders had better overviews of what is funded across instruments and individual research councils, charities, and research foundations, they would have an easier time targeting areas with too little support.

Third, private funders will naturally focus their funding allocations. When excellence becomes an overarching quality criteria across the system, topics with little market potential are at a disadvantage. Funding from private sources spill over in publicly funded research because it bolsters the amount of resources available to high performing groups. The results presented here at least hints at this type of spill-over as the funded topics and disciplines across funding sectors coincide. Other research have also shown that the most successful researchers often attract funding from diverse sources, and support their research with both public and private funds (Madsen & Aagaard, 2020). Research strains with success in private funding will then inadvertently become successful in public funding competition. To counter this, we need greater coordination between funders. This could be through shared priority setting that ensures some degree of functional division of labor between research funding bodies, and continuous coordination in e.g. a national research funding forum.

From my analysis, it is impossible to argue for any causal link from individual concentration to topic concentration, or from thematic priorities to topic concentration.

Or rather, we could credibly argue that the connection runs in both directions. Given the complexity of the interlinked and interdependent mechanisms at work, it is rather likely that funding concentration and research priorities are prone to mutually reinforcing cycles. This does not necessarily invalidate the results of the analysis. The great degree of overlap mandates more research. Future studies could investigate how researchers change topic along with priorities and prestige. Here, we would need in-depth case studies to tease apart the causal mechanism at work, especially to take into account the timing of different events.

10.6 Supplementary material

10.6.1 Comparison of grant samples

Table S10.1: Summary of grant samples for full and publication-matched samples

	Denmark		United Kingdom	
	Full sample	WoS-matched sample	Full sample	WoS-matched sample
No. of grants	18,404	11,605	41,416.0	30,662.0
No. of PIs	8,624	4,703	22,290.0	14,465.0
Mean amount[a]	2,555,962	2,731,366	434,732.0	586,523.5
Median amount[a]	984,191	1,000,000	239,807.2	333,547.9
Std. dev. of amount[a]	6,690,068	7,103,392	1,425,052.0	1,745,664.5

Note: [a] Values are given in DKK and CPI-adjusted pounds respectively.

10.6.2 Formal definitions of models for priority overlap

To assess the degree of overlap in priorities between both targeted vs. non-targeted and private vs. public funding, I use regression models to quantify both the average degree of overlap, the corresponding uncertainty in this average, and the remaining variation in the data. For figure 10.4 and 10.6, the models are simple linear regressions, while supplementary analyses are hierarchical linear models. The formal definitions of these models follow here.

Simple overall models. Let y denote the calculated funding percentile in non-targeted funding instruments (for figure 10.4) or public research councils (for figure 10.6), and i the index for individual topics (micro or meso) or discipline (reference-based or journal-based). Then:

$$y_i \sim Normal(\mu_i, \sigma)$$

The expression for μ_i specify the average funding percentile in non-targeted/public funding as a linear function of the funding percentile in targeted/private instruments for all topics or disciplines i:

$$\mu_i = \alpha_0 + \beta_1 P_i$$

The parameter β_i then represents the average difference in funding percentile between an average topic and an average topic with a one point higher funding percentile. To recover posterior distributions of the parameters α_0, β_1, and σ, we define their distributions through the following priors:

$$\alpha_0 \sim Normal(50, 20)$$

$$\beta_1 \sim Normal(0.5, 1)$$

$$\sigma \sim Half\text{-}normal(0, 20)$$

These priors allow for funding percentile overlap to be both negatively and positively related, with 95 % of parameter values between - 1.48 and 2.42 (see coloured area in figure S10.1. Because Madsen and Aagaard (2020) find a positive relationship between funding percentiles between public and private funders for disease related research, the prior on β_1 puts around 70 % of probability mass above zero, and it thus induces some skepticism towards negative relationships between funding percentiles.

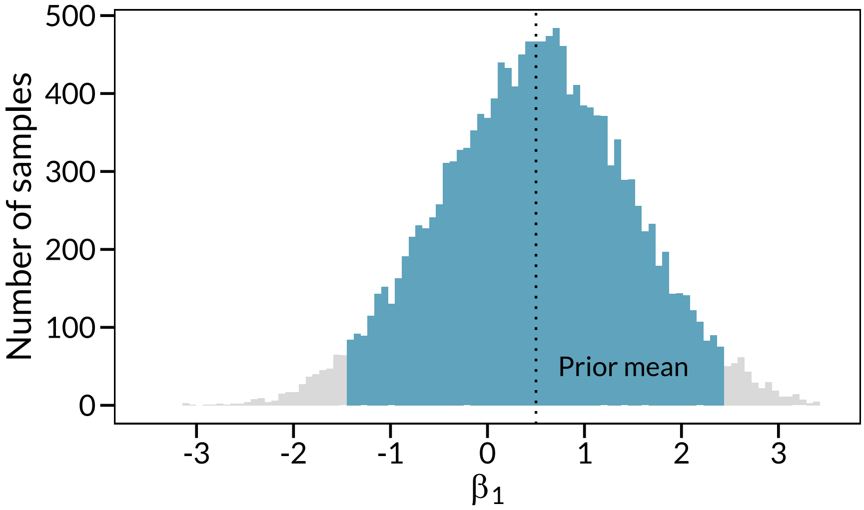


Figure S10.1: Prior distribution for priority overlap (β_1).

Random effects on research area. Because thematic priorities are often applied unequally across research areas, and private funders in Denmark are focused on the biomedical and natural sciences, I account for this difference by estimating a hierarchical model, where the degree of overlap in percentiles, i.e. β_1 and α_0, can vary between the five research areas. The model builds on the basic bivariate model from above, but adds some parameters:

$$y_i \sim Normal(\mu_i, \sigma)$$

$$\mu_i = \alpha_0 + \alpha_j + (\beta_0 + \beta_J)P_{ij}$$

Here, α_0 is the average intercept across five areas j, α_j is the varying offsets from the global intercept, β_0 is the average degree of overlap, while β_j is the offsets from this for each research area j. For both α_0 and β_0, the priors are identical to the ones presented above. I add priors for the standard deviations among intercepts and β-coefficients, and for the correlation matrix **R** of all the varying intercepts and effects:

$$\sigma_\alpha \sim Normal(0, 20)$$

$$\sigma_\beta \sim Normal(0, 1)$$

$$\mathbf{R} \sim LKJ(2)$$

where the LKJ prior on is just a weakly informative regularising prior skeptical of extreme correlations.

Estimating the models. All posterior distributions from the models were estimated empirically by Hamiltonian Markov Chain Monte Carlo (HMC) using R (ver. 3.6.3) and the package 'brms' (Bürkner, 2017, 2018). For all models, I assigned 1,000 burn-in samples and 5,000 iterations for the sampler, using four chains, an acceptance rate of 0.99 (adapt delta), and a max tree depth of 15. The visualisations of the posterior distributions and the posterior predictions were made with the R packages 'tidybayes' (Kay, 2019) and 'ggplot2' (Wickham, 2016).

10.6.3 Results from hierarchical models of priority overlap

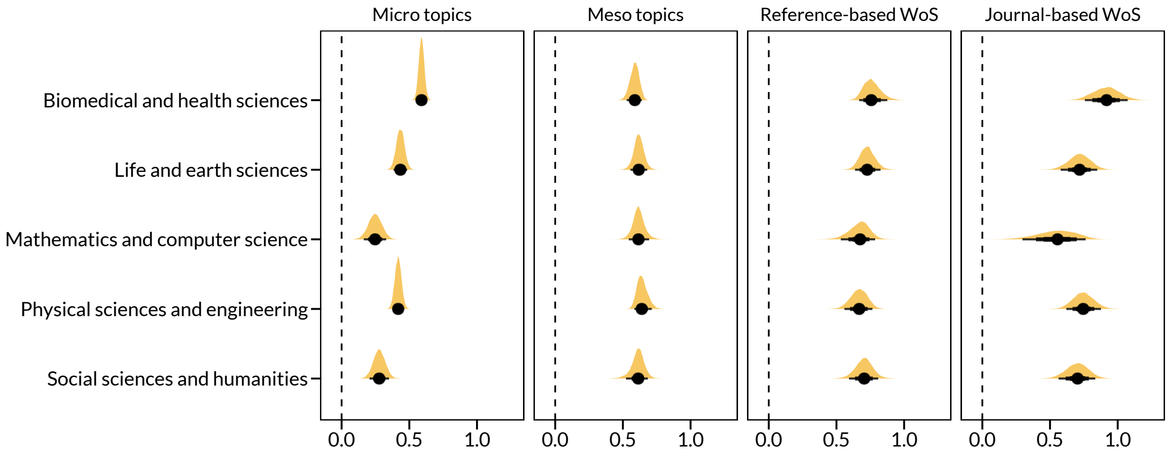


Figure S10.2: Topical overlap of funded strategic and "free" research in Denmark across five research areas. *Note:* Points are posterior medians, with 50 %, 70 % and 90 % credible intervals calculated from the posterior distribution.

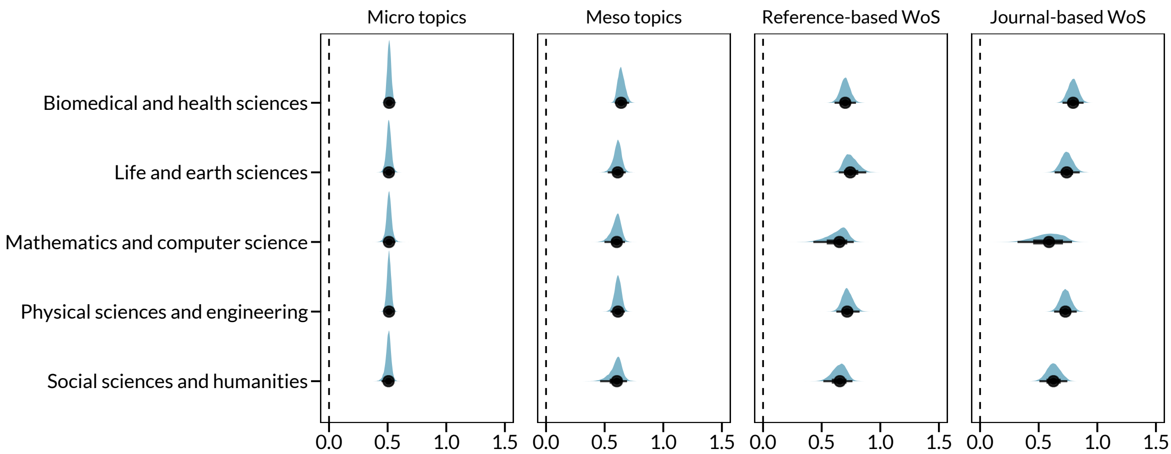


Figure S10.3: Topical overlap of funded strategic and "free" research in the UK across five research areas. *Note:* Points are posterior medians, with 50 %, 70 % and 90 % credible intervals calculated from the posterior distribution.

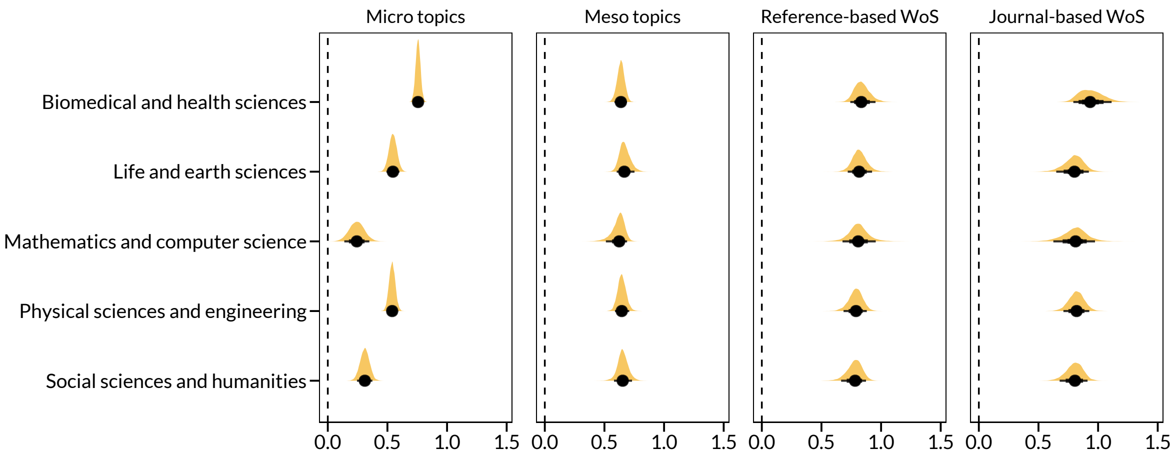


Figure S10.4: Topical overlap of funded private vs. public research in Denmark across five research areas. *Note:* Points are posterior medians, with 50 %, 70 % and 90 % credible intervals calculated from the posterior distribution.

References

Aagaard, K. (2011). *Kampen om basismidlerne: historisk institutionel analyse af basisbevillingsmodellens udvikling på universitetsområdet i Danmark*. Aarhus Universitet. https://books.google.dk/books?id=tJM2twAACAAJ

Aagaard, K. (2017). The Evolution of a National Research Funding System: Transformative Change Through Layering and Displacement. *Minerva*, *55*(3), 279–297.

Aagaard, K. (2019). *Quality criteria and concentration of research funding*, Nordic Institute for Studies in Innovation, Research and Education. https://www.r-quest.no/wp-content/uploads/2019/12/R-Quest%7B%5C_%7DPolicy%7B%5C_%7Dbrief%7B%5C_%7D4%7B%5C_%7D2019.pdf

Aagaard, K., Kladakis, A. & Nielsen, M. W. (2020). Concentration or dispersal of research funding? *Quantitative Science Studies*, *1*(1), 1–33.

Aagaard, K., Schneider, J. W. & Andersen, J. P. (2019). *Koncentration eller spredning? Fordeling af konkurrenceudsatte forskningsmidler på tværs af danske bevillingsorganer* (tech. rep.). Center for Forskningsanalyse, Aarhus Universitet. Aarhus.

Agar, J., Clarke, S., Craig, C., Edgerton, D. & Flanagan, K. (2019). *Lessons from the History of UK Science Policy* (tech. rep. August). The British Academy. London. https://www.thebritishacademy.ac.uk/documents/243/Lessons-History-UK-science-policy.pdf

Ahlgren, P., Chen, Y., Colliander, C. & van Eck, N. J. (2020). Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. *Quantitative Science Studies*, *1*(2), 714–729.

Alberts, B., Kirschner, M. W., Tilghman, S. & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences*, *111*(16), 5773–5777.

Aleta, A., Meloni, S. & Perra, N. (2019). Explore with caution: Mapping the evolution of scientific interest in physics. *EPJ Data Science*, *28*(27), 1–15.

Allison, P. D., Long, J. S. & Krauze, T. K. (1982). Cumulative Advantage and Inequality in Science. *American Sociological Review*, *47*(5), 615–625.

Allison, P. D. & Stewart, J. A. (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review*, *39*(4), 596–606.

Atkinson, A. B., Piketty, T. & Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, *49*(1), 3–71.

Azoulay, P., Graff Zivin, J. S., Li, D. & Sampat, B. N. (2019). Public R&D Investments and Private-sector Patenting : Evidence from NIH Funding. *Review of Economic Studies*, *86*(1), 117–152.

Baldwin, M. (2018). Scientific Autonomy, Public Accountability, and the Rise of "Peer Review" in the Cold War United States. *Isis*, *109*(3), 538–558.

BBSRC. (2020). Responsive mode priorities. https://bbsrc.ukri.org/funding/grants/priorities/

Berg, J. M. (2012). Well-funded investigators should receive extra scrutiny. *Nature*, *489*(7415), 203.

Berg, J. M. (2013). Scientific approaches to science policy. *Molecular Biology of the Cell*, *24*(21), 3273–3274.

Best, R. K. (2012a). Disease Politics and Medical Research Funding. *American Sociological Review*, *77*(5), 780–803.

Best, R. K. (2012b). *Illnesses as Interests: The Rise of Disease Advocacy and the Politics of Medical Research* (Doctoral dissertation). University of California, Berkeley. http://escholarship.org/uc/item/0vp5b83r

Blei, D. M. & Lafferty, J. D. (2006). Dynamic Topic Models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120.

Blei, D. M. & Lafferty, J. D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, *1*(1), 17–35.

Bloch, C., Graversen, E. K. & Pedersen, H. S. (2014). Competitive Research Grants and Their Impact on Career Performance. *Minerva*, *52*(1), 77–96.

Bloch, C., Schneider, J. W. & Sinkjær, T. (2016). Size, Accumulation and Performance for Research Grants : Examining the Role of Size for Centres of Excellence. *PLoS ONE*, *11*(2), e0147726.

Bloch, C. & Sørensen, M. P. (2015). The size of research funding: Trends and implications. *Science and Public Policy*, *42*(1), 30–43.

Bol, T., de Vaan, M. & van de Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences*, 1–4.

Bonaccorsi, A. & Daraio, C. (2005). Exploring size and agglomeration effects on public research productivity. *Scientometrics*, *63*(1), 87–120.

Bornmann, L. (2017). Measuring impact in research evaluations: a thorough discussion of methods for, effects of and problems with impact measurements. *Higher Education*, *73*(5), 775–787.

Bornmann, L. & Daniel, H. D. (2006). Selecting scientific excellence through committee peer review - A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics, 68*(3), 427–440.

Bornmann, L., De Moya Anegón, F. & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS ONE, 5*(10), e13327.

Bornmann, L. & Mutz, R. (2015). Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References. *Journal of the American Society for Information Science and Technology, 66*(11), 2215–2222.

Bornmann, L., Wallon, G. & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. *PLoS ONE, 3*(10), e3480.

Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information, 14*(6), 19–47.

Bowen, A. & Casadevall, A. (2015). Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. *Proceedings of the National Academy of Sciences, 112*(36), 11335–11340.

Boyack, K. W. & Klavans, R. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351–374.

Boyack, K. W., Klavans, R., Small, H. & Ungar, L. (2014). Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science. *Journal of Engineering and Technology Management, 32,* 147–159.

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., Schijvenaars, B., Skupin, A., Ma, N. & Börner, K. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS ONE, 6*(3), e18029.

Boyack, K. W., Smith, C. & Klavans, R. (2018). Toward predicting research proposal success. *Scientometrics, 114*(2), 449–461.

Bozeman, B. (2020). Public Value Science. *Issues in Science and Technology, 36*(4), 34–41.

Bozeman, B. & Sarewitz, D. (2011). Public Value Mapping and Science Policy Evaluation. *Minerva, 49*(1), 1–23.

Brattström, E. & Hellström, T. (2019). Street-level priority-setting: The role of discretion in implementation of research, development, and innovation priorities. *Energy Policy, 127*(December 2018), 240–247.

Braun, D. (1998). The role of funding agencies in the cognitive development of science. *Research Policy, 27*(8), 807–821.

Bromham, L., Dinnage, R. & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature, 534*(7609), 684–687.

Bruggeman, J., Traag, V. A. & Uitermark, J. (2012). Detecting Communities through Network Data. *American Sociological Review*, *77*(6), 1050–1063.

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411.

Bush, V. (1945). *Science, the Endless Frontier. A Report to the President* (tech. rep.). US Government Print Office. Washington, DC.

Carter, C. F. (1963). The Distribution of Scientific Effort. *Minerva*, *1*(2), 172–181.

Cassi, L., Lahatte, A., Ràfols, I., Sautier, P. & de Turckheim, É. (2017). Improving fitness: Mapping research priorities against societal needs on obesity. *Journal of Informetrics*, *11*(4), 1095–1113.

Chubin, D. E. (1976). The Conceptualization of Scientific Specialties. *The Sociological Quaterly*, *17*(4), 448–476.

Chubin, D. E. (1985). Beyond Invisible Colleges: Inspirations and Aspirations of Post-1972 Social Studies of Science. *Scientometrics*, *7*(3-6), 221–254.

Ciarli, T. & Ràfols, I. (2019). The relation between research priorities and societal demands: The case of rice. *Research Policy*, *48*(4), 949–967.

Cobham, A., Schlögl, L. & Sumner, A. (2016). Inequality and the Tails : the Palma Proposition and Ratio. *Global Policy*, *7*(1), 25–36.

Cole, J. R. & Cole, S. (1972). The Ortega Hypothesis. *Science*, *178*(4059), 368–375.

Cole, J. R. & Zuckerman, H. (1975). The emergence of a scientific specialty: The self-exemplifying case of the sociology of science. In L. A. Coser (Ed.), *The idea of social structure: Papers in honor of robert k. merton* (pp. 139–174). Harcourt Brace Jovanovich.

Cole, S. (1970). Professional Standing and the Reception of Scientific Discoveries. *American Journal of Sociology*, *76*(2), 286–306.

Cole, S. & Cole, J. R. (1967). Scientific Output and Recognition: A Study in the Operation of the Reward System in Science. *American Sociological Review*, *32*(3), 377–390.

Cole, S., Cole, J. R. & Simon, G. A. (1981). Chance and Consensus in Peer Review. *Science*, *214*(4523), 881–886.

Collins, H. M. (1985). The Possibilities of Science Policy. *Social Studies of Science*, *15*(3), 554–558.

CWTS. (2020). CWTS Leiden Ranking: Fields. Retrieved August 31, 2020, from https://www.leidenranking.com/information/fields%7B%5C#%7Dmicro-level-fields

Danthi, N. S., Wu, C. O., DiMichele, D. M., Keith Hoots, W. & Lauer, M. S. (2015). Citation impact of NHLBI R01 grants funded through the American recovery and reinvestment act as compared to R01 grants funded through a standard payline. *Circulation Research*, *116*(5), 784–788.

Dasgupta, P. & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*(5), 487–521.

DiPrete, T. A. & Eirich, G. M. (2006). Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments. *Annual Review of Sociology*, *32*(1), 271–297.

Domenico, M. D., Omodei, E. & Arenas, A. (2016). Quantifying the diaspora of knowledge in the last century. *Applied Network Science*, *1*(15), 1–13.

Ebadi, A. & Schiffauerova, A. (2015). How to receive more funding for your research? get connected to the right people. *PLoS ONE*, *10*(7), 1–19.

Edwards, A. M., Isserlin, R., Bader, G. D., Frye, S. V., Wilson, T. M. & Yu, F. H. (2011). Too many roads not taken. *Science*, *470*(7333), 163–165.

Evans, J. A. (2010). *Industry collaboration, scientific sharing, and the dissemination of knowledge* (Vol. 40).

Evans, J. A., Shim, J. M. & Ioannidis, J. P. (2014). Attention to local health burden and the global disparity of health research. *PLoS ONE*, *9*(4), 1–9.

Fang, F. C., Bowen, A. & Casadevall, A. (2016). NIH peer review percentile scores are poorly predictive of grant productivity. *eLife*, *5*(e13323), 1–6.

Fang, F. C. & Casadevall, A. (2016). Research Funding: the Case for a Modified Lottery. *mBio*, *7*(2), e00422–16.

Fedderke, J. W. & Goldschmidt, M. (2015). Does massive funding support of researchers work?: Evaluating the impact of the South African research chair funding initiative. *Research Policy*, *44*(2), 467–482.

Fortin, J. M. & Currie, D. J. (2013). Big Science vs. Little Science: How Scientific Impact Scales with Funding. *PLoS ONE*, *8*(6).

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. & Barabási, A. L. (2018). Science of science. *Science*, *359*(6379).

Foster, J. G., Rzhetsky, A. & Evans, J. A. (2015). Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, *80*(5), 875–908.

Franssen, T., Scholten, W., Hessels, L. K. & de Rijcke, S. (2018). The Drawbacks of Project Funding for Epistemic Innovation: Comparing Institutional Affordances and Constraints of Different Types of Research Funding. *Minerva*, *56*(1), 11–33.

Frickel, S., Gibbon, S., Howard, J., Kempner, J., Ottinger, G. & Hess, D. J. (2010). Undone science: Charting social movement and civil society challenges to research agenda setting. *Science Technology and Human Values*, *35*(4), 444–473.

Gallo, S. A., Carpenter, A. S., Irwin, D., McPartland, C. D., Travis, J., Reynders, S., Thompson, L. A. & Glisson, S. R. (2014). The validation of peer review through research impact measures and the implications for funding strategies. *PLoS ONE, 9*(9), e106474.

Gallo, S. A. & Glisson, S. R. (2018). External Tests of Peer Review Validity Via Impact Measures. *Frontiers in Research Metrics and Analytics, 3*(22), 1–7.

Garfield, E. (1978). High impact science and the case of Arthur Jensen. *Current Comments, 41*, 652–662.

Gazni, A., Sugimoto, C. R. & Didegah, F. (2012). Mapping World Scientific Collaboration: Authors, Institutions, and Countries. *Journal of the American Society for Information Science and Technology, 63*(2), 323–335.

Gerow, A., Hu, Y., Boyd-Graber, J., Blei, D. M. & Evans, J. A. (2018). Measuring discursive influence across scholarship. *Proceedings of the National Academy of Sciences, 115*(13), 3308–3313.

Geuna, A. (2001). The Changing Rationale for European University Research Funding: Are There Negative Unintended Consequences? *Journal of Economic Issues, 35*(3), 607–632.

Geuna, A. & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*(4), 277–304.

Gieryn, T. F. (1978). Problem retention and problem change in science. In J. Gaston (Ed.), *Sociology of science* (pp. 96–115). Jossey-Bass.

Gillum, L. A., Gouveia, C., Dorsey, E. R., Pletcher, M., Mathers, C. D., Charles, E. & Johnston, S. C. (2011). NIH Disease Funding Levels and Burden of Disease. *PLoS ONE, 6*(2), e16837.

Gläser, J. (2019). How can governance change research content? Linking science policy studies to the sociology of science. In D. Simon, S. Kuhlmann, J. Stamm & W. Canzler (Eds.), *Handbook on science and public policy* (pp. 419–447). Edward Elgar Publishing.

Gläser, J., Glänzel, W. & Scharnhorst, A. (2017). Same data — different results ? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics, 111*(2), 981–998.

Gläser, J. & Laudel, G. (2016). Governing Science. *European Journal of Sociology, 57*(1), 117–168.

Gläser, J. & Velarde, K. S. (2018). Changing Funding Arrangements and the Production of Scientific Knowledge: Introduction to the Special. *Minerva, 56*(1), 1–10.

Gök, A., Rigby, J. & Shapira, P. (2015). The Impact of Research Funding on Scientific Outputs: Evidence From Six Smaller European Countries. *Journal of the American Society for Information Science and Technology, 67*(3), 715–730.

Goldstone, J. A. (1979). A Deductive Explanation of the Matthew Effect in Science. *Social Studies of Science, 9*(3), 385–391.

Greisdorf, H. & O'Connor, B. (2003). Nodes of Topicality: Modeling User Notions of on Topic Documents. *Journal of the American Society for Information Science and Technology, 54*(14), 1296–1304.

Griffiths, T. L. & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*(suppl 1), 5228–5235.

Gross, C., Anderson, G. & Powe, N. (1999). The Relation between Funding by the National Institutes of Health and the Burden of Disease. *New England Journal of Medicine, 340*(24), 1881–1887.

Gross, K. & Bergstrom, C. T. (2018). Contest models highlight inherent inefficiencies of scientific funding competitions. *PLoS Biology, 17*(1), e3000065.

Gross, K. & Bergstrom, C. T. (2021). *Why ex post peer review encourages high-risk research while ex ante review discourages it*. http://arxiv.org/abs/2106.13282

Hagstrom, W. O. (1970). Factors related to the use of different modes of publishing research in four scientific fields. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 85–124). Lexington Books.

Hagstrom, W. O. (1974). Competition in Science. *American Sociological Review, 39*(1), 1–18.

Hallonsten, O. & Hugander, O. (2014). Supporting 'future research leaders' in Sweden: Institutional isomorphism and inadvertent funding agglomeration. *Research Evaluation, 23*(3), 249–260.

Hammarfelt, B. (2020). Discipline. *Knowledge Organization, 47*(3), 244–256.

Hand, E. (2012). Extra scrutiny for ' grandee grantees'. *Nature, 482*(7386), 450–451.

Hand, E. & Wadman, M. (2008). 222 NIH grants: 22 Researchers. *Nature, 452*(7185), 258.

Hanna, M. (2015). Matching taxpayer funding to population health needs. *Circulation Research, 116*(8), 1296–1300.

Hargens, L. L., Mullins, N. C. & Hecht, P. K. (1980). Research Areas and Stratification Processes in Science. *Social Studies of Science, 10*, 55–74.

Head, M. G., Fitchett, J. R., Cooke, M. K., Wurie, F. B., Hayward, A. C. & Atun, R. (2013). UK investments in global infectious disease research 1997-2010: A case study. *The Lancet Infectious Diseases, 13*(1), 55–64.

Head, M. G., Fitchett, J. R., Moore, D. A. & Atun, R. (2015). Systematic analysis of funding awarded to institutions in the United Kingdom for infectious disease research, 1997–2010. *JRSM Open, 6*(3), 205427041557705.

Head, M. G., Fitchett, J. R., Nageshwaran, V., Kumari, N., Hayward, A. & Atun, R. (2016). Research Investments in Global Health: A Systematic Analysis of UK Infectious

Disease Research Funding and Global Health Metrics, 1997-2013. *EBioMedicine*, *3*, 180–190.

Hegde, D. (2009). Political influence behind the veil of peer review : an analysis of public biomedical research funding in the U.S . *Journal of Law and Economics*, *52*(4), 665–690.

Hegde, D. & Mowery, D. C. (2008). Politics and Funding in the U.S. Public Biomedical R&D System. *Science*, *322*(December), 1797–1798.

Hegde, D. & Sampat, B. (2015). Can Private Money Buy Public Science? Disease Group Lobbying and Federal Funding for Biomedical Research. *Management Science*, *61*(10), 2281–2298.

Heinze, T. (2008). How to sponsor ground-breaking research: A comparison of funding schemes. *Science and Public Policy*, *35*(5), 302–318.

Held, M., Laudel, G. & Gläser, J. (2021). Challenges to the validity of topic reconstruction. *Scientometrics*, *126*(5), 4511–4536.

Held, M. & Velden, T. (2019). How to interpret algorithmically constructed topical structures of research specialties? A case study comparing an internal and an external mapping of the topical structure of invasion biology. *17th International Conference on Scientometrics and Informetrics, ISSI 2019 - Proceedings*, *2*, 1933–1939.

Hellström, T. & Jacob, M. (2012). 'Weingberg's Choice': Classic Tensions in the Concept of Scientific Merit. *Minerva*, *50*(3), 381–396.

Hellström, T., Jacob, M. & Sjöö, K. (2017). From thematic to organizational prioritization: The challenges of implementing RDI priorities. *Science and Public Policy*, *44*(5), 599–608.

Hendricks, M. (2019). Men get bigger science grants than women in Canada. Retrieved March 2, 2021, from https://michael-hendricks.medium.com/men-get-bigger-science-grants-in-canada-6033da63d832

Hernandez-Villafuerte, K., Sussex, J., Robin, E., Guthrie, S. & Wooding, S. (2017). Economies of scale and scope in publicly funded biomedical and health research: Evidence from the literature. *Health Research Policy and Systems*, *15*(1), 1–17.

Hess, D. J. (2009). The potentials and limitations of Civil Society Research: Getting undone science done. *Sociological Inquiry*, *79*(3), 306–327.

Hess, D. J. (2015). Beyond Scientific Consensus: Scientific Counterpublics, Countervailing Industries, and Competing Research Agendas. In W. Viehöver, P. Wehling & S. Koenen (Eds.), *The public shaping of medical research: Patient associations, health movements, and biomedicine* (pp. 151–171). Routledge.

Hicks, D. & Katz, J. S. (2011). Equity and Excellence in Research Funding. *Minerva*, *49*(2), 137–151.

Hjørland, B. (1992). The concept of 'subject' in information science. *Journal of Documentation*, *48*(2), 172–200.

Hjørland, B. (2001). Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content... and Relevance. *Journal of the American Society for Information Science and Technology*, *52*(9), 774–778.

Hjørland, B. (2017). Reviews of concepts in knowledge organization: Subject (documents). *Knowledge Organization*, *44*(1), 55–64.

Hjørland, B. & Nicolaisen, J. (2005). Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". In F. Crestani & I. Ruthven (Eds.), *5th international conference on conceptions of library and information sciences, colis 2005, glasgow, uk june 4-8, 2005, proceedings* (pp. 96–106). Lecture Notes in Computer Science.

HM Treasury. (2003). *Lambert Review of Business-Univeristy Collaboration* (tech. rep. December). HM Treasury. London. www.lambertreview.org.uk%20For

Ho, D. E., Imai, K., King, G. & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*(3), 199–236.

Hoonlor, A., Szymanski, B. K. & Zaki, M. J. (2013). Trends in computer science research. *Communications of the ACM*, *56*(10), 74.

Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., Hutchins, B. I., Davis, A. F., Lauer, M. S., Valantine, H. A., Anderson, J. M. & Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Science Advances*, *5*(10), eaaw7238.

Horlings, E. & Gurney, T. (2013). Search strategies along the academic lifecycle. *Scientometrics*, *94*(3), 1137–1160.

Horta, H. & Santos, J. M. (2020). Organisational factors and academic research agendas: an analysis of academics in the social sciences. *Studies in Higher Education*, *45*(12), 2382–2397.

Hutchins, W. J. (1978). The Concept of Aboutness in Subject Indexing. *Aslib Proceedings*, *30*(5), 172–181.

Iacus, S. M., King, G. & Porro, G. (2009). cem: Software for Coarsened Exact Matching. *Journal of Statistical Software*, *30*(9), 1–27.

Iacus, S. M., King, G. & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361.

Iacus, S. M., King, G. & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24.

Ida, T. & Fukuzawa, N. (2013). Effects of large-scale research funding programs: a Japanese case study. *Scientometrics*, *94*(3), 1253–1273.

Ioannidis, J. P. A. (2011). More time for research: Fund people not projects. *Nature, 477*(7366), 529–531.

Jacob, B. A. & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of Public Economics*, *95*(9-10), 1168–1177.

Jia, T., Wang, D. & Szymanski, B. K. (2017). Quantifying patterns of research-interest evolution. *Nature Human Behaviour*, *1*(4), 1–7.

Jones, R. & Wilsdon, J. (2018). *The Biomedical Bubble: Why UK research and innovation needs a greater diversity of priorities, politics, places and people*. Nesta.

Katz, Y. & Matter, U. (2020). Metrics of Inequality: The Concentration of Resources in the U .S. Biomedical Elite. *Science as Culture*, *29*(4), 475–502.

Kay, M. (2019). tidybayes: Tidy Data and Geoms for Bayesian Models.

Kearnes, M. & Wienroth, M. (2011). Tools of the Trade: UK Research Intermediaries and the Politics of Impacts. *Minerva*, *49*(2), 153–174.

Klavans, R. & Boyack, K. W. (2011). Using Global Mapping to Create More Accurate Document-Level Maps of Research Fields. *Journal of the American Society for Information Science and Technology*, *62*(1), 1–18.

Klavans, R. & Boyack, K. W. (2017a). Research portfolio analysis and topic prominence. *Journal of Informetrics*, *11*(4), 1158–1174.

Klavans, R. & Boyack, K. W. (2017b). The research focus of nations: Economic vs. altruistic motivations. *PLoS ONE*, *12*(1), e0169383.

Klavans, R. & Boyack, K. W. (2017c). Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? *Journal of the American Society for Information Science and Technology*, *68*(4), 984–998.

Krishnan, A. (2009). *What are academic disciplines?* (Tech. rep. No. 03/09). National Center for Research Methods.

Krumhansl, J. A. & Trigg, G. L. (1975). Indexing and Classification. *Physical Review Letters*, *34*(17), 1065–1066.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd ed.). Univeristy of Chicago Press.

Kuhn, T. S. (1977). *The Essential Tension: Selected Studies in Scientific Tradition and Change*. University of Chicago Press.

Kundu, O. & Matthews, N. E. (2019). The role of charitable funding in university research. *Science and Public Policy*, *46*(4), 611–619.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A. & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, *46*(3), 417–435.

Larivière, V., Macaluso, B., Archambault, É. & Gingras, Y. (2010). Which scientific elites? On the concentration of research funds, publications and citations. *Research Evaluation, 19*(1), 45–53.

Laudel, G. (2006a). The 'quality myth': Promoting and hindering conditions for acquiring research funds. *Higher Education, 52*(3), 375–403.

Laudel, G. (2006b). The art of getting funded: How scientists adapt to their funding conditions. *Science and Public Policy, 33*(7), 489–504.

Laudel, G. & Gläser, J. (2014). Beyond breakthrough research: Epistemic properties of research and their consequences for research funding. *Research Policy, 43*(7), 1204–1216.

Lauer, M., Roychowdhury, D., Patel, K., Walsh, R. & Pearson, K. (2017). *Marginal returns and levels of research grant support among scientists supported by the national institutes of health* (tech. rep.).

Lauer, M. S., Danthi, N. S., Kaltman, J. & Wu, C. (2015). Predicting Productivity Returns on Investment: Thirty Years of Peer Review, Grant Funding, and Publication of Highly Cited Papers at the National Heart, Lung, and Blood Institute. *Circulation Research, 117*(3), 239–243.

Lauer, M. S., Doyle, J., Wang, J. & Roychowdhury, D. (2021). Associations of topic-specific peer review outcomes and institute and center award rates with funding disparities at the national institutes of health. *eLife, 10*(e67173), e67173.

Leisyte, L. & Dee, J. R. (2012). Understanding Academic Work in a Changing Institutional Environment. In J. Smart & M. Paulsen (Eds.), *Higher education: Handbook of theory and research* (vol. 25, pp. 123–206). Springer.

Leisyte, L., Enders, J. & Boer, H. D. (2010). Mediating Problem Choice: Academic Researchers ' Responses to Changes in their Institutional Environment. In R. Whitley, J. Gläser & L. Engwall (Eds.), *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 266–290). Oxford University Press.

Lerchenmueller, M. J. & Sorenson, O. (2018). The gender gap in early career transitions in the life sciences. *Research Policy, 47*(6), 1007–1017.

Leydesdorff, L. & Bornmann, L. (2016). The Operationalization of "Fields" as WoS Subject Categories (WCs) in Evaluative Bibliometrics: The Cases of "Library and Information Science" and "Science & Technology Studies". *Journal of the American Society for Information Science, 67*(3), 707–714.

Li, D. & Agha, L. (2015). Big names or big ideas: Do peer-review panels select the best science proposals? *Science, 348*(6233), 434–438.

Li, D., Azoulay, P. & Sampat, B. N. (2017). The applied value of public investments in biomedical research: Supplement. *Science, 356*(6333), 78–81.

Li, J., Xie, Y., Wu, D. & Chen, Y. (2017). Underestimating or overestimating the distribution inequality of research funding? The influence of funding sources and subdivision. *Scientometrics, 112*(1), 55–74.

Lind, J. K., Hernes, H., Pulkkinen, K. & Söderlind, J. (2019). External Research Funding and Authority Relations. In R. Pinheiro, L. Geschwind, H. Foss Hansen & K. Pulkkinen (Eds.), *Reforms, organizational change and performance in higher education: A comparative account from the nordic countries* (pp. 145–180). Palgrave Macmillan.

Long, J. S. (1978). Productivity and Academic Position in the Scientific Career. *American Sociological Review, 43*(6), 889–908.

Lorsch, J. R. (2015). Maximizing the return on taxpayers ' investments in fundamental biomedical research. *Molecular Biology of the Cell, 26*(9), 1578–1582.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences, 16*(10), 317–323.

Luukkonen, T. & Thomas, D. A. (2016). The 'Negotiated Space' of University Researchers' Pursuit of a Research Agenda. *Minerva, 54*(1), 99–127.

Ma, A., Mondragón, R. J. & Latora, V. (2015). Anatomy of funded research in science. *Proceedings of the National Academy of Sciences, 112*(48), 14760–14765.

MacRoberts, M. H. & MacRoberts, B. R. (1986). Quantitative Measures of Communication in Science: A Study of the Formal Level. *Social Studies of Science, 16*(1), 151–172.

MacRoberts, M. H. & MacRoberts, B. R. (1987). Testing the Ortega hypothesis: Facts and artifacts. *Scientometrics, 12*(5-6), 293–295.

MacRoberts, M. H. & MacRoberts, B. R. (2010). Problems of Citation Analysis: A Study of Uncited and Seldom-Cited Influence. *Journal of the American Society for Information Science and Technology, 61*(1), 1–13.

Madsen, E. B. & Aagaard, K. (2020). Concentration of Danish research funding on individual researchers and research topics: Patterns and potential drivers. *Quantitative Science Studies, 1*(3), 1159–1181.

Manton, K. G., Gu, X.-L., Lowrimore, G., Ullian, A. & Tolley, H. D. (2009). NIH funding trajectories and their correlations with US health dynamics from 1950 to 2004. *Proceedings of the National Academy of Sciences, 106*(27), 10981–10986.

Martin, B. R. & Irvine, J. (1985). Evaluating the Evaluators: A Reply to Our Critics. *Social Studies of Science, 15*(3), 558–575.

McMahan, P. & McFarland, D. A. (2021). Creative Destruction: The Structural Consequences of Scientific Curation. *American Sociological Review, 86*(2), 341–376.

Merton, R. K. (1957). Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review, 22*(6), 635–659.

Merton, R. K. (1968). The Matthew Effect in Science. *Science, 159*(3810), 56–63.
Merton, R. K. (1988). The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property. *Isis, 79*(4), 606–623.
Milojevic, S. (2014). Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences, 111*(11), 3984–3989.
Milojević, S. (2010). Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. *Journal of the American Society for Information Science, 61*(7), 1410–1423.
Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative Science Studies, 1*(1), 183–206.
Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science and Technology, 48*(9), 810–832.
Mongeon, P., Brodeur, C., Beaudry, C. & Lariviere, V. (2016). Concentration of research funding leads to decreasing marginal returns. *Research Evaluation, 25*(4), 396–404.
Mongeon, P. & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics, 106*(1), 213–228.
Morris, S. A. & Van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology, 42*, 213–295.
Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics, 12*(4), 103–134.
Nag, S., Yang, H., Buccola, S. & Ervin, D. (2013). Productivity and financial support in academic bioscience. *Applied Economics, 45*(19), 2817–2826.
Nature. (2003). Editorial: Research funding: The problem with priorities. *Nature Materials, 2*(10), 639.
Nelson, R. R. (1959). The Simple Economics of Basic Research. *Journal of Political Economy, 67*(3), 297–306.
Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics, 100*(3), 741–754.
Nicolaisen, J. (2002). The J-shaped distribution of citedness. *Journal of Documentation, 58*(4), 383–395.
Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*(1), 609–641.
Nuffield Council on Bioethics. (2014). *The Culture of Scientific Research in the UK* (tech. rep. December). https://www.nuffieldbioethics.org/wp-content/uploads/Nuffield%7B%5C_%7Dresearch%7B%5C_%7Dculture%7B%5C_%7Dfull%7B%5C_%7Dreport%7B%5C_%7Dweb.pdf

Park, H., Lee, J. & Kim, B. C. (2015). Project selection in NIH: A natural experiment from ARRA. *Research Policy*, *44*(6), 1145–1159.

Peifer, M. (2017). The argument for diversifying the NIH grant portfolio. *Molecular Biology of the Cell*, *28*(22), 2935–2940.

Polanyi, M. (1962). The Republic of Science: Its Political and Economic Theory. *Minerva*, *1*(1), 54–73.

Price, D. J. d. S. (1963). *Little Science, Big Science*. Columbia University Press.

Price, D. J. d. S. (1965). Networks of Scientific Papers. *Science*, *149*(3683), 510–515.

Price, D. J. d. S. (1976). A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science and Technology*, *27*(5), 292–306.

Price, D. J. d. S. (1986). *Little Science, Big Science... and Beyond*. Columbia University Press.

Ràfols, I., Hopkins, M. M., Hoekman, J., Siepel, J., O'Hare, A., Perianes-Rodríguez, A. & Nightingale, P. (2014). Big Pharma, little science? A bibliometric perspective on Big Pharma's R&D decline. *Technological Forecasting and Social Change*, *81*(1), 22–38.

Ràfols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *Research Policy*, *41*(7), 1262–1282.

Ràfols, I. & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287.

Ràfols, I., Porter, A. L. & Leydesdorff, L. (2010). Science Overlay Maps: A New Tool for Research Policy and Library Management. *Journal of the American Society for Information Science*, *61*(9), 1871–1887.

Ràfols, I. & Yegros-Yegros, A. (2018). Is research responding to health needs? *Social Observatory of the La Caixa Foundation*, 1–13. https://observatoriosociallacaixa.org/en/-/responde-la-investigacion-a-las-necesidades-de-salud

Reardon, S. (2014). Lobbying sways NIH grants. *Nature*, *515*, 19.

Reskin, B. F. (1977). Scientific Productivity and the Reward Structure of Science. *American Sociological Review*, *42*(3), 491–504.

Rip, A. (1994). The republic of science in the 1990s. *Higher Education*, *28*(1), 3–23.

Rogers, J. D., Youtie, J. & Kay, L. (2012). Program-level assessment of research centers: Contribution of Nanoscale Science and Engineering Centers to US Nanotechnology National Initiative goals. *Research Evaluation*, *21*(5), 368–380.

Rotolo, D., Ràfols, I., Hopkins, M. M. & Leydesdorff, L. (2015). Strategic Intelligence on Emerging Technologies: Scientometric Overlay Mapping. *Journal of the American Society for Information Science and Technology*, *68*(1), 214–233.

Røttingen, J.-A., Regmi, S., Eide, M., Young, A. J., Viergever, R. F., Årdal, C., Guzman, J., Edwards, D., Matlin, S. A. & Terry, R. F. (2013). Mapping of available health research and development data : what ' s there , what ' s missing , and what role is there for a global observatory ? *The Lancet*, *382*(9900), 1286–1307.

Rzhetsky, A., Foster, J. G., Foster, I. T. & Evans, J. A. (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences*, *112*(47), 14569–14574.

Sampat, B. N. (2012). Mission-oriented biomedical research at the NIH. *Research Policy*, *41*(10), 1729–1741.

Sarewitz, D. (2016). Saving Science. *The New Atlantis*, *49*(Spring/Summer), 4–40.

Sarewitz, D. & Pielke, R. A. (2007). The neglected heart of science policy: reconciling supply of and demand for science. *Environmental Science and Policy*, *10*(1), 5–16.

Schamber, L., Eisenberg, M. B. & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management*, *26*(6), 755–776.

Schneider, J. W., Horbach, S. P. & Aagaard, K. (2021). Stop blaming external factors: A historical-sociological argument. *Social Science Information*, *60*(3), 1–9.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, *59*(3), 425–465.

Sjögårde, P. & Ahlgren, P. (2018). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of topics. *Journal of Informetrics*, *12*(1), 133–152.

Sjögårde, P. & Ahlgren, P. (2020). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of specialties. *Quantitative Science Studies*, *1*(1), 207–238.

Sjögårde, P., Ahlgren, P. & Waltman, L. (2021). Algorithmic labeling in hierarchical classifications of publications: Evaluation of bibliographic fields and term weighting approaches. *Journal of the Association for Information Science and Technology*, *72*(7), 853–869.

Small, H. (1980). Co-citation content analysis and the structure of paradigms. *Journal of Documentation*, *36*(3), 183–196.

Small, H. & Griffith, B. C. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialties. *Science Studies*, *4*(1), 17–40.

Smith, A. (2019). From PACS to PhySH. *Nature Reviews Physics*, *1*(1), 8–11.

Sörlin, S. (2007). Funding Diversity: Performance-based Funding Regimes as Drivers of Differentiation in Higher Education Systems. *Higher Education Policy*, *20*(4), 413–440.

Stage, A. K. & Aagaard, K. (2020). National policies as drivers of organizational change in universities: A string of reinforcing reforms. *Quantitative Science Studies*, *1*(2), 849–871.

Statistics Denmark. (2020). FOUOFF07: R&D-expenses in public sector by subject and type of expense. Retrieved November 12, 2020, from https://www.statbank.dk/10148

Stephan, P. E. (2012). Perverse incentives. *Nature*, *484*(5), 29–31.

Stewart, J. (1995). Models of priority-setting for public sector research. *Research Policy*, *24*(1), 115–126.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719.

Stoeger, T., Gerlach, M., Morimoto, R. I. & Nunes Amaral, L. A. (2018). Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS Biology*, *16*(9), 1–25.

Stoeger, T. & Nunes Amaral, L. A. (2020). COVID-19 research risks ignoring important host genes due to pre-established research patterns. *eLife*, *9*, e61981.

Strategic Research Council. (2014). *Strategisk forskning, 2004-2014* (tech. rep.). Styrelsen for Forskning og Innovation. Copenhagen. https://ufm.dk/publikationer/2014/filer-2014/arsskrift%7B%5C_%7D2013-14%7B%5C_%7Ddk.pdf

Sugimoto, C. R., Ahn, Y. Y., Smith, E., Macaluso, B. & Larivière, V. (2019). Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. *The Lancet*, *393*(10171), 550–559.

Sugimoto, C. R. & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation*, *71*(4), 775–794.

The Danish Council for Research and Innovation Policy. (2020a). *Forskningsfinansiering – forandringer og konsekvenser* (tech. rep.). Ministry of Higher Education and Science. Copenhagen.

The Danish Council for Research and Innovation Policy. (2020b). *Forskningsfinansiering i Danmark – hvad er der i vente?* (Tech. rep.). Ministry of Higher Education and Science. Copenhagen.

Toulmin, S. (1963). The Complexity of Scientific Choice: A Stocktaking. *Minerva*, *1*(2), 343–359.

Traag, V. A., Waltman, L. & Eck, N. J. V. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, *9*(5233), 1–12.

Turner, S. P. & Chubin, D. E. (1976). Another appraisal of Ortega, the Coles, and science policy: the Ecclesiastes hypothesis. *Social Science Information*, *15*(4/5), 657–662.

Vaesen, K. & Katzav, J. (2017). How much would each researcher receive if competitive government research funding were distributed equally among researchers ? *PLoS ONE*, *12*(9), e0183967.

van den Besselaar, P. & Heimeriks, G. (2001). Disciplinary, multidisciplinary, interdisciplinary – concepts and indicators. *8th Conference on Scientometrics and Informetrics, presented at the ISSI2001, Sydney*.

Wadman, M. (2010). Study says middle sized labs do best. *Nature*, *468*(7322), 356–357.

Wahls, W. P. (2016). Biases in grant proposal success rates , funding rates and award sizes affect the geographical distribution of funding for biomedical research. *PeerJ*, *4*, e1917.

Wahls, W. P. (2018a). *High cost of bias: Diminishing marginal returns on NIH grant funding to institutions* (tech. rep.). bioRxiv.

Wahls, W. P. (2018b). The NIH must reduce disparities in funding to maximize its return on investments from taxpayers. *eLife*, *7*, e34965.

Wahls, W. P. (2019). The National Institutes of Health needs to better balance funding distributions among US institutions. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(27), 13150–13154.

Wallace, M. L. & Ràfols, I. (2015). Research Portfolio Analysis in Science Policy: Moving from Financial Returns to Societal Benefits. *Minerva*, *53*(2), 89–115.

Wallace, M. L. & Ràfols, I. (2018). Institutional shaping of research priorities: A case study on avian influenza. *Research Policy*, *47*(10), 1975–1989.

Waltman, L., Boyack, K. W., Colavizza, G. & van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, *1*(2), 691–713.

Waltman, L., Ràfols, I., Eck, N. J. V. & Yegros, A. (2019). *Supporting priority setting in science using research funding landscapes* (No. 1), Research on Research Institute. https://doi.org/10.6084/m9.figshare.9917825.v1

Waltman, L. & Van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B*, *86*(11), 471.

Waltman, L. & van Eck, N. (2012). A New Methodology for Constructing a Publication-Level Classification System of Science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2396.

Wang, J., Lee, Y. N. & Walsh, J. P. (2018). Funding model and creativity in science: Competitive versus block funding and status contingency effects. *Research Policy*, *47*(6), 1070–1083.

Weinberg, A. M. (1963). Criteria for Scientific Choice. *Minerva*, *1*(2), 159–171.

Whitley, R. (1970). Black Boxism and the Sociology of Science: A Discussion of the Major Developments in the Field. *The Sociological Review*, *18*(1), 61–92.

Whitley, R., Gläser, J. & Laudel, G. (2018). The Impact of Changing Funding and Authority Relationships on Scientific Innovations. *Minerva*, *56*(1), 109–134.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wu, J. (2015). Distributions of scientific funding across universities and research disciplines. *Journal of Informetrics*, *9*(1), 183–196.

Wuchty, S., Jones, B. F. & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(May), 1036–1039.

Yan, E. (2014). Research dynamics: Measuring the continuity and popularity of research topics. *Journal of Informetrics*, *8*(1), 98–110.

Yan, E., Ding, Y. & Jacob, E. K. (2012). Overlaying communities and topics: an analysis on publication networks. *Scientometrics*, *90*(2), 499–513.

Yao, L., Li, Y., Ghosh, S., Evans, J. A. & Rzhetsky, A. (2015). Health ROI as a measure of misalignment of biomedical needs and resources. *Nature Biotechnology*, *33*(8), 807–811.

Yegros-Yegros, A., Klippe, W. V. D., Abad-garcia, M. F. & Ràfols, I. (2020). Exploring why global health needs are unmet by research efforts: the potential influences of geography , industry and publication incentives. *Health Research Policy and Systems*, *18*(47), 1–14.

Yegros-Yegros, A., Tjissen, R., Abad-Garcia, M.-F. & Ràfols, I. (2018). Drug research priorities at odds with global disease toll. *Nature Index*, *16 March*. https://www.natureindex.com/news-blog/drug-research-priorities-at-odds-with-global-disease-toll

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E. & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature Communications*, *10*(3439), 1–11.

Zhi, Q. & Meng, T. (2016). Funding allocation, inequality, and scientific research output: an empirical study based on the life science sector of Natural Science Foundation of China. *Scientometrics*, *106*(2), 603–628.

Ziman, J. H. (1987). The problem of ”problem choice”. *Minerva*, *25*(1), 92–106.

Zuckerman, H. (1967). Nobel Laureates in Science : Patterns of Productivity , Collaboration , and Authorship. *American Sociological Review*, *32*(3), 391–403.

Zuckerman, H. (1970). Stratification in American Science. *Sociological Inquiry*, *40*(2), 235–257.

Zuckerman, H. (1978). Theory choice and problem choice in science. *Sociological Inquiry*, *48*(3-4), 65–95.

Zuckerman, H. (1992). The proliferation of Prizes: Nobel Complements and Nobel Surrogates in the Reward System of Science. *Theoretical Medicine*, *13*, 217–231.

Summary

Prior to the Second World War, governments were scantly involved in the funding of research. Massive investments in military-related research during the war changed this, and continued into the post-war period with large research programmes in defence, nuclear energy, and space technology. Along with new investments in higher education and Ph.D.-programmes, the new era of government-sponsored research signalled a pronounced change in the scientific enterprise. Over the course of the 20th century, the scientific system would expand gradually, employ an increasing amount of researchers, and command a larger and larger budget.

During the past 40 years, renewed interest in different ways of funding research have emerged as public research budgets have stagnated and interests in prioritising resources have reached the political level. The most important development have unequivocally been the turn to competitive funding allocations as a method of stimulating increased research performance and efficient use of scarce resources. A dual-streamed funding system with floor funding and block grants to universities, coupled with project grants offered through competition between individual researchers, have now become the norm in most countries. The hope has been for competitive funding schemes to further knowledge generation through the selection of top researchers and research environments, promote specific research fields, and strengthen research quality through internal competition between scientists and more external cooperation outside the university.

Throughout the 1980s, 1990s, and 2000s the trend towards more reliance on competitive funding have strengthened, with larger shares of the budget being offered through grant competitions and more targeted investment through mission-oriented or strategic funding programmes. Private research funders have also taken up the gauntlet and begun offering more funding opportunities based on application and peer-review. This turn towards a more competitive allocation of research funding may be a necessary step in planning and directing research to the benefit of economic growth and societal needs, while also rationalising resources. On the other hand, observers ascribe many of the problems faced by modern science to exactly this development. Competitive research funding have been criticised for skewing resources towards a too narrow

set of perspectives and people, hampering free scientific development and innovation, and leaving researchers with uncertain career prospects. Most agree that some concentration of resources is desirable to reward and incentivise the best researchers and studies on the most promising topics, disciplines, and research fields. The question is how far competitive research funding have pushed the degree of concentration and how the structure of competitive funding systems have added to this?

In this dissertation, I attempt to assess how funding is distributed across different categories of research: topics, specialities, diseases, disciplines, and research areas. This is a descriptive effort in investigating how concentrated funding is in terms of research content, instead of focusing on who receives funding.

To do this, I employ newly collected data on around 120,000 research grants awarded by 22 research-funding organisations to researchers in Denmark and the United Kingdom from 2005 through 2016. These are uncommonly detailed data on research funding, which track who received a grant at what time from who. However, they tell us very little about the research content produced within the scope of each individual grant. Instead, the data is enriched with information about scholarly publications produced by the grantees, with a view to characterising the thematic components of each grant through these. Research content can be categorised at many different levels of granularity from the very detailed topics to broadly descriptive disciplines. It turns out that – no matter the level of analysis - competitively awarded funding is strongly concentrated. Grants flow to a small group of disciplines within the natural and biomedical sciences, but strong stratification also takes place within fields of science. Molecular biology and clinical medicine are prioritised to a much higher degree than plant science or public health, and within clinical medicine, diabetes research receives much more funding than research in chronic pulmonary disease.

The strong degree of concentration across all categories of research content is likely driven by multiple interacting factors. One aspect is path dependence and positive feedback of research funding. Some topics and disciplines receive much more funding than others, and keep attracting funding as time goes by. Well-funded researchers also tend to accumulate even more funding, and when most researchers seldom change the topics they study, individual funding concentration turns into topical concentration. Another factor pertains to the intended prioritisation of certain disciplines and topics by research funders. When private and public funders support the same researchers working with a narrow problem area, it leads to a more skewed allocation of resources than intended. Certain researchers are more likely to receive funding because the study prioritised topics and disciplines, which in turn amplify strong concentration. The introduction of prioritised funding by both governmental and private funders have led to a spill-over of priorities to the rest of the funding system. This dissertation has taken the first steps in more thoroughly understanding what type of research is being funded through competitive means, and argues that if we are to mitigate a lack of diversity in research approaches, we need policies that view the competitive funding landscape from a topical portfolio perspective.

Dansk resumé

Før Anden Verdenskrig stod forskningsfinansiering ikke højt på dagsordenen hos nationale regeringer. Massive investeringer i militærrelateret forskning under krigen medvirkede dog til at ændre denne prioritering, og nye forskningsprogrammer inden for forsvarsteknologi, atomenergi og rumteknologi forblev en fast del af efterkrigstidens forskningspolitik. Sammen med øgede investeringer i universitets- og ph.d.-uddannelse, signalerede en øget grad af offentligt finansieret forskning, at systemet nu undergik afgørende forandringer. Over de følgende 70 år har forskningssystemet ekspanderet kraftigt med et støt stigende antal forskere og større og større finansieringskrav.

De seneste års stagnerende forskningsbudgetter og større behov for strategisk prioritering af ressourcer har medført en fornyet interesse for forskellige måder at finansiere forskning på. Den vigtigste politiske udvikling har her været introduktionen af konkurrenceudsatte forskningsmidler som en metode til at sikre forskningskvalitet og en effektiv fordeling af knappe ressourcer. Et to-strenget finansieringssystem, hvor basisbevillinger til universiteter komplimenteres med projektbevillinger udbudt i åben konkurrence, er nu normen i langt de fleste lande. Forestillingen har længe været, at konkurrencebaseret fordeling af forskningsmidler skulle fremme vidensproduktionen ved at identificere de dygtigste forskere og forskningsmiljøer, fremme bestemte lovende forskningsfelter, og styrke forskningskvaliteten gennem øget intern konkurrence og eksternt samarbejde. Op gennem 1980'erne, 1990'erne og 2000'erne har de konkurrencebaserede midler optaget en større og større andel af det totale forskningsbudget, og flere midler uddeles gennem strategiske programmer vendt mod bestemte forskningsområder. Private finansieringsorganer og fonde har fulgt denne tendens, og uddeler flere og flere forskningsmidler baseret på åbne ansøgningsrunder og fagfællebedømmelse.

For nogle har denne udvikling været et nødvendigt skridt i forsøget på at styre forskningsressourcerne mod centrale samfundsbehov og sikre fremtidig økonomisk vækst. For andre har udviklingen bidraget til flere problemer end den løser. Den øgede brug af konkurrenceudsættelse er blevet kritiseret for at koncentrere forskningsmidler på for få hænder, og favorisere et snævert udsnit af forskningsområder. Stigende konkurrence stækker den frie forskning, skaber usikre karriereveje for unge forskere og mindsker den tid forskere har til deres kerneopgaver. De fleste er dog enige om, at nogen grad af koncentration er nødvendig for belønne og fremme de bedste forskere, samt de mest lovende forskningsemner- og områder. Spørgsmålet er hvor langt brugen af konkurrenceudsættelse har presset denne koncentration, og hvordan uddelingskriterierne har biddraget til dette.

Jeg forsøger I denne afhandling at afdække graden af koncentration i de konkurrencebaserede forskningsmidler og fordelingen af disse på tværs af forskningsemner og discipliner. Modsat tidligere undersøgelser fokuserer jeg her på indholdet af forskningen, frem for hvem der modtager midlerne. I undersøgelsen benytter jeg nyindsamlet data om ca. 120.000 individuelle forskningsbevillinger fra 22 forskellige forskningsfinansierende organisationer i Danmark og Storbritannien i perioden 2005 til 2017. Dette data repræsenterer en hidtil uset detaljegrad, og indeholder oplysninger om hvilke forskere har modtaget midler fra både offentlige og private forskningsråd og –fonde i en tolvårig periode. Disse oplysninger fortæller os dog i mindre grad hvilken type af forskning er bedrevet for eksterne midler. Jeg kombinerer derfor disse bevillingsdata med oplysninger om forskningsartikler udgivet af hver individuel bevillingsmodtager, med henblik på at karakterisere det tematiske indhold i hver forskningsbevilling.

Dette indhold kan karakteriseres og grupperes på flere måder og med varierende detaljegrad fra brede disciplinære områder til specifikke forskningsemner. Uanset analyseniveau viser det sig dog, at de konkurrenceudsatte forskningsmidler er særdeles skævt fordelte. Bevillingerne koncentreres i overvejende grad på en lille gruppe af natur- og sundhedsvidenskabelige discipliner, men også inden for disse discipliner er der stor ulighed. Molekylærbiologi og klinisk medicin prioriteres i langt højere grad end plante- eller folkesundhedsvidenskab, mens forskere inden for klinisk medicin modtager langt flere midler, hvis de forsker i diabetes end kronisk obstruktiv lungesyndrom (KOL).

Den stærke koncentration på tværs af discipliner og emner skyldes flere interagerende faktorer. Et aspekt er den udprægede stiafhængighed i forskningsfinansieringen, og de positive feedbackmekanismer det skaber. Nogle emner og discipliner modtager lang flere midler end andre, og denne forskel består eller sågar stiger over tid. Højt finansierede forskere akkumulerer også langt flere midler end deres mindre succesfulde kollegaer, og når forskere tilmed sjældent skifter mellem forskningsemner øges både den individuelle og indholdsmæssige koncentration. En anden faktor omhandler bevillingsgivernes prioritering af bestemte emner. Når både private og offentlige forskningsfonde støtter en mindre andel af succesfulde forskere, i bestemte forskningsmiljøer, skævvrides fordelingen af midler ud over det forventelige. Bestemte forskere har flere finansieringsmuligheder og måske højere succesrate, fordi de studerer bestemte emner, hvilket medvirker til en stigende koncentration. Jeg har i denne afhandling taget de første skridt i retningen mode en mere dybdegående forståelse af hvilken type forskning finansieres gennem konkurrenceudsatte forskningsmidler. For at modvirke en faldende emnemæssig diversitet i forskningen kræves der et målrette fokus på forskningspolitik, der i højere grad rettes mod hvad vi finansierer end hvem vi finansierer.