Innovation Under Resource Constraints: Supercomputing in Scientific Research

Justine Boudou* and John McKeon+

October 17, 2024

.

Abstract

Resource constraints significantly shape innovation by influencing both the rate and direction of output. In this paper, we study how constraints of a key input to frontier innovation, high-performance computing power, impact innovative output. Our empirical setting is an NSF-funded initiative called XSEDE, which allows researchers to apply for access to a distributed network of high-performance computing resources, or "supercomputers," across the United States. Leveraging the unique resource allocation process of XSEDE, we are able to isolate the impact of computing resource constraints on scientific production. We find that an increase in the share of computing resources allocated to a researcher relative to what they require (i.e., a relaxing of constraints) leads to a large increase in the number of papers published and alters the research direction. Specifically, additional computing resources lead researchers to study less popular and newer topics, explore new topics that they have not studied in their prior work, and broaden the scope of their projects. Our findings underscore the intricate relationship between resources and innovative output, with consequences for optimizing resource allocation strategies and fostering innovation.

Keywords: *Innovation*, *Science*, *Resource Allocation*, *Constraints*, *Computing Resources*

Authors contributed equally, listed alphabetically. *Boudou: Harvard Business School (jboudou@hbs.edu) +McKeon: Boston University (jmckeon@bu.edu)

1 Introduction

Resource constraints exert a significant influence on the rate and direction of innovation (Arrow 1962, Nelson 1959, Barney 1991, Agrawal et al. 2018). Constraints can impact firms' risk-taking (Krieger et al. 2022) and R&D investments (Mezzanotti & Simcoe 2023), along with investors' and firms' tolerance for failure (Nanda & Rhodes-Kropf 2017, Atanassov 2016). Prior work has shown that access to new resources, such as technology and data, impacts innovation by shaping researchers' effort allocation, collaboration patterns, and the types of projects pursued (Furman & Teodoridis 2020, Ding et al. 2010, Nagaraj & Tranchero 2023). However, we still need a better understanding of how variation in the amount of resources impacts innovation at the individual level, particularly when it comes to the ability of researchers to expand the boundaries of their research and push the knowledge frontier. This is important to understand given that individuals are often the focal point of resource allocation decisions made by policymakers and R&D managers (Babina et al. 2023, Boudreau et al. 2016, Blomfield & Vakili 2023) who operate in environments with limited resources. These decisions are particularly crucial when they involve resources that may enable researchers to pursue cutting-edge, less conventional work. This has been difficult to study because variation in resource constraints is usually difficult to observe, coincides with variation in other resources, or is subject to a selection process where more promising projects receive more resources.

This paper offers a novel empirical setting that allows us to study how constraints on a critical input to innovation, supercomputing power, impact the innovative output of scientific researchers. After researchers submit applications for supercomputing resources to support projects tackling complex, computationally intensive problems, these requests are evaluated and reduced based on the overall demand and available supply at the time. This creates variation in the extent to which researchers are more or less constrained relative to their initial requests so that we can estimate the impact of resource constraints on the rate and type of research output. Importantly, our data enables us to observe the quantity of resources a researcher objectively requires for their project, allowing us to account for the underlying projects' scale and potential that is rarely observable in other settings.

Beyond these empirical advantages, studying computing resources is crucial given that they are a key input to modern innovation and the production of frontier knowledge (Thompson et al. 2022, Kim 2023, Tranchero 2023). Across many fields, access to high-performance computing ("HPC") resources - often referred to as "supercomputers" - has become indispensable for firms' competitive advantage and for researchers striving to push the boundaries of knowledge. In the field of artificial intelligence for instance, firms increasingly rely on massive supercomputing clusters to support the training of their prediction models.[1] Advanced computational resources facilitate the screening of potential drug candidates and the simulation of complex weather patterns, accelerating the development of new therapies and aiding in disaster preparedness. Despite its widespread importance, computing power remains a scarce resource for many innovators (Lohr 2019). In light of this, recent policy discussions in the U.S. and the E.U. emphasize the importance of additional investment in HPC resources for national competitiveness, firms' competitive advantage in the global market, and national security.[2] These inequalities in resource availability underline the importance of understanding innovators' choices when their computing power is constrained.

We first present a conceptual framework to motivate our empirical analysis, which considers the impact of supercomputing resource constraints on the rate and type of researchers' output. Intuitively, one might expect that as resource constraints are relaxed, researchers will produce additional output. However, the effect of increased high-performance computing resources on the choice of research topic - and whether researchers shift towards more innovative (or "less conventional") work - is less clear. For instance, if researchers do not act strategically based on these constraints, there would be no change in a given researcher's choice of topic. On the other hand, given the heterogeneous nature of innovative projects and the incentives researchers face, researchers may prioritize certain projects over others as resource constraints are relaxed, thereby altering the characteristics of the output produced (Azoulay et al. 2011). Thus, how computing constraints shape the direction of innovation and in particular researchers' ability to engage in less conventional work is an open question. As suggested by prior literature, we consider how variations in computing resources may affect researchers' choice to pursue less popular topics, research lines they haven't worked on before, and projects broader in scope.

We empirically examine how variations in supercomputing resource constraints affect researchers' innovative output in the novel context of an NSF-funded initiative known as *The Extreme Science and Engineering Discovery Environment* (henceforth "XSEDE"). XSEDE allows researchers to apply for access to a distributed network of high-performance computing resources, or "supercomputers," across the United States. XSEDE supported over $6 billion of funded research over a ten year period, playing a pivotal role in enabling researchers to tackle computationally intensive scientific challenges that were previously beyond reach.[3] Our dataset

[1]For example, Microsoft has recently discussed investing $100 billion to build a new supercomputer, "Stargate." See https://fortune.com/2024/04/02/microsoft-openai-stargate-100-billion-ai-supercomputer-star-wars-sdi/

[2]See report from the NSA-DOE https://www.nitrd.gov/nitrdgroups/images/b/b4/nsa_doe_hpc_techmeetingreport.pdf, the think-thank ITIF https://www2.itif.org/2016-high-performance-computing.pdf, and the European Investment Bank https://www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf

[3]See NSF Awards 1053575 and 1548562.

combines internal XSEDE data on 711 projects and their computing allocations with information on the publication output associated with each project retrieved from the Dimensions database.

In order to examine our research question, we exploit the allocation process of XSEDE which provides us with plausibly exogenous variation in how constrained researchers are in the amount of computing resources they receive. Academic researchers seeking access via XSEDE first request a computing allocation for their project, which is assessed by a committee. The committee provides a recommended allocation based solely on the amount of computing resources the project *would need* if XSEDE resources were unconstrained. Thus, we are able to account for the amount of resources a project objectively requires, helping us to control for its underlying quality or scale. Throughout the period we study, demand for computing resources consistently exceeds supply. Hence, on a quarterly basis, the aggregate computing demand is then reconciled with available supply by reducing each recommendation based on an algebraic formula and allocating researchers this amount. Our empirical framework leverages this unique resource allocation process of XSEDE in order to create variation in the quantity of computing resources received by researchers, hence how constrained they are.

We find that increased resource allocation leads to a large increase in the number of papers published within a project. Specifically, a 10% increase in allocation leads to a 15% increase in the number of papers written. On average, we find that as resource constraints are relaxed, researchers tend to explore less studied (i.e., "unconventional") research avenues, as measured by a decrease in the use of research topics studied by other researchers. Further, researchers receiving more computing resources tend to broaden the scope of their research, pursuing projects which span a wider array of topics. Delving deeper, we find that these effects are driven by the *least* conventional and *broadest* papers, respectively. We also find that researchers explore new research streams among their *most* exploratory work, studying topics they had not previously studied. In addition, among the papers addressing the *newest* topics, researchers are increasingly focusing on even more recent topics. Overall, as resource constraints are relaxed, papers that address less common topics tend to explore even more niche areas, papers focusing on recent ideas delve into even newer ones, the most exploratory projects become more so, and the broadest papers expand their scope further. These results suggest that additional computing resources enable researchers to push the boundaries of their work, advancing into different and less conventional territories. They do this by studying newer topics that have not yet gained traction in the scientific community, working in areas they have not previously worked on, and creating broader papers. We also find evidence that increased resource allocation leads to a decrease in the number of citations but has no impact on journal quality, suggesting the papers *may* struggle to accrue citations due to the topics they study rather than their inherent quality.

Taken together, our study sheds light on the intricate relationship between resource allocation and innovative output, underscoring the importance of effective resource allocation strategies among managers and policymakers in promoting innovation. We emphasize the importance of understanding the changes in the attributes of output produced when the amount of resources varies. At a high level, our findings underscore the significance of considering not only the rate but also the direction of research output when evaluating resource allocation strategies at R&D intensive firms or publicly funded programs that allocate scarce resources across individuals.

This paper makes three distinct contributions. First, we contribute to the literature studying the impact of resource constraints on innovation by shedding light on the impact of constraints at the individual level. Much of this literature focuses on the effect of funding and financial constraints on firms' innovation output (Hoegl et al. 2008), showing for example that financing impacts firms' quantity and quality of innovations (Atanassov 2016), the type of firms that investors are willing to fund (Nanda & Rhodes-Kropf 2017, Conti et al. 2019) as well as the composition of their R&D investments (Mezzanotti & Simcoe 2023). Our contribution shifts the focus to the individual level. We examine how resource constraints influence the decisions that individual researchers make, providing a more granular view of how innovators adjust their efforts and outputs in response to varying resource conditions. Additionally, our analysis complements work on other non-fungible constraints such as geography (Singh & Marx 2013, Catalini 2018), policy (Glennon 2023, Conti 2018) or environmental regulation (Berrone et al. 2013) on innovation, as well as the role of other scarce inputs such as attention (Fini et al. 2022).

Second, we contribute to the related literature on the marginal returns to resource allocation. Prior work studies the returns of R&D subsidies directly on firms' patenting and revenue (Bronzini & Piselli 2016, Howell 2017) or through the spillovers they generate (Azoulay et al. 2019, Myers & Lanahan 2022). In the same vein, our empirical setting allows us to explore the *intensive margin* of computing resource allocation. This is usually hard to estimate given that variation for access to non-fungible resources usually comes from shocks that vary *whether* firms or individuals get access or not to inputs or technology (Kong et al. 2022, Jacob & Lefgren 2011, Nagaraj & Tranchero 2023, Furman & Teodoridis 2020, Ding et al. 2010) but not necessarily *how much* they get access to. In our context, it is crucial to understand this margin as the relevant constraint to pursuing frontier research is not necessarily whether a researcher has access to *any* computing power, but rather *how much* computing power they have access to.

Third, we contribute to the literature on firms' and researchers' innovation strategies which explores risk-taking (Mandler 2017, Krieger et al. 2022, Franzoni & Stephan 2023), effort allocation (Blomfield & Vakili

2023) and other changes in innovation direction such as the departure from one's existing line of research (Myers 2020, Nagle & Teodoridis 2020) or the expansion of existing projects (Myers & Tham 2023). Building upon this literature, our study investigates how high-performance computing resource allocation influences researchers' innovation strategies. By analyzing the impact of increased resource allocation on research output and research direction, we offer insights into the mechanisms driving innovation within the context of advanced computing resources. Our findings provide valuable contributions to understanding the complex interplay between resource allocation decisions, researchers' strategic choices, and the pursuit of innovative and cutting-edge scientific research.

Section 2 lays out a simple conceptual framework to ground our empirics. Section 3 describes the XSEDE setting and Section 4 discusses our empirical approach. Section 5 describes our data, Section 6 outlines our findings, and Section 7 concludes.

2 Researcher Strategies Under Resource Constraints

2.1 Do researchers produce more output as resource constraints are relaxed?

We begin by examining how an increase in the amount of computing resources allocated to researchers (i.e., the relaxing of constraints) impacts the quantity of output they produce. In practice, it is not obvious that more resources necessarily leads to an increase in output. Researchers could potentially be less efficient with additional resources, or they might leverage external alternatives, which could offset the expected gains. However, based on previous work and the limited availability of outside options, we might expect that an increase in the amount of computing resources leads to an increase in output.

First, existing literature tends to find a positive correlation between the availability of resources and scientific or innovation-related outputs, typically measured by publications or patents. For instance, access to data or to technology such as Information Technology (IT) is associated with an increase in the number of publications (Ding et al. 2010, Nagaraj & Tranchero 2023). Mezzanotti & Simcoe (2023) indicates that investments in basic research decline when funding cost increases. Additional work shows that tax credit stimulates additional R&D activity by reducing the cost of R&D (e.g., Hall & Van Reenen 2000). In the scientific context, research suggests that government funding leads to increased research output (Payne et al. 1999, Arora & Gambardella 2005). These insights indicate that enhancing resource availability generally leads to greater research productivity.

Importantly, prior work suggests that the impact of getting more resources on research output largely depends on the availability of outside options. For instance, Jacob & Lefgren (2011) find that being funded by the National Institute of Health (NIH) has a positive but only small effect on the number of publications and citations and show that this is consistent with researchers being able to shift to another source of funding. In contrast, high-performance computing resources typically lack direct substitutes, due to their specialized nature and the significant infrastructure required to support them (Dongarra et al. 2011, Millett & Fuller 2011). Indeed, unlike other research inputs that are fungible, HPC systems are uniquely designed to handle large-scale simulations, complex data analysis, and computationally intensive tasks. Further, large simulations often cannot be split up onto multiple resources but requires an allocation within one environment. Hence, outside options are limited. As a result, we expect that an increase in the amount of computing resources will have a positive impact on researchers' output.

2.2 Do researchers produce a different type of output as resource constraints are relaxed?

While we establish that the quantity of output is likely to increase with additional computing resources, the type of innovation produced when resource constraints are relaxed might also differ. First, prior literature suggests that technological, spatial, institutional, and fungible constraints have meaningful consequences on innovation characteristics (Catalini 2018, Boyabatlı et al. 2016, Cerqueiro et al. 2017) and highlights mechanisms such as risk-aversion, competition, or task reallocation (Krieger et al. 2022, Mezzanotti & Simcoe 2023, Furman & Teodoridis 2020). Since computing resources represent a critical input for researchers and are subject to availability constraints, an increase in these resources could similarly impact the nature of the science pursued. Moreover, additional computing resources may make researchers more tolerant of the career risk associated with undertaking more uncertain or less "conventionally incentivized" projects (e.g., Azoulay et al. 2011). Indeed, while incentives in science may lead researchers to prioritize projects and activities that are more likely to advance their careers (Stephan 1996, Azoulay et al. 2007), additional resources may increase researchers' willingness to take on projects with different characteristics (Manso 2011). Overall, the specific way in which computing resources shape the type of science pursued by researchers remains under-explored and warrants further investigation. We examine three potential avenues that researchers may reasonably pursue with more resources: (i) exploring less-studied research avenues (ii) venturing beyond their expertise (iii) broadening their research scope. While these strategies are not exhaustive, they offer a view into crucial ways in which computing resources might influence the type of science researchers engage in.

First, researchers may decide to use additional computing resources to pursue less-studied research avenues, either by studying topics that are inherently less popular or by focusing on topics that are younger and have appeared more recently in the scientific literature. The prevailing norms and incentives within the scientific community tend to prioritize the pursuit of topics that yield tangible returns, such as papers and citations (Dasgupta & David 1994). While studying less-studied areas might allow researchers to establish priority on a specific topic, it may also come at the cost of having a smaller audience and making the peer-review process more difficult (Krieger et al. 2023), ultimately increasing the perceived risk of pursuing these topics. As suggested by Azoulay et al. (2011), incentives such as the tolerance for early failure can lead researchers to change their research direction and to choose less-traveled scientific avenues. By reducing researchers' constraints getting additional computing resources may similarly allow researchers to bear more risk and study less-traveled research avenues.

Second, the relaxing of resource constraints may lead researchers to embark on research endeavors that are 'new-to-them' and to venture beyond their existing areas of expertise (Hill et al. 2021). Along these lines, a large body of innovation literature studies the exploration-exploitation trade-off, its underlying mechanisms, as well as its boundary conditions (Luger et al. 2018, Marino et al. 2015, Danneels & Sethi 2011, Sidhu et al. 2007, Choi et al. 2016, Fitzgerald et al. 2021, Zhuo 2022). While various incentives in the modern scientific community tend to encourage specialization (Jones 2009, Foster et al. 2015), researchers may be more willing to pursue projects outside their specialty - which are inherently more uncertain - as constraints are relaxed.[4] Hence, an increase in computing resources might increase researchers' exploration of new topics that are not part of their existing knowledge base.

Third, researchers may broaden their scope by pursuing more interdisciplinary or cross-cutting topics that traverse traditional disciplinary boundaries. Breadth, or scope, and its relationship with firm performance or innovation has been studied in various streams of literature, such as organizational learning (Dutt & Lawrence 2022, Ferguson & Carnabuci 2017). While prior research often uses knowledge breadth as a mechanism that impacts innovation (Nagle & Teodoridis 2020, Zhou & Li 2012), the scope of research may also be endogenous to the constraints researchers face. However, our understanding of how resource constraints impact the breadth of researchers' projects remains limited. Because scientific norms tend to reward researchers' expertise on a specific piece of knowledge (Jones 2009), researchers might not have strong incentives to pursue broad projects at first. In addition, broader projects might structurally require more resources to get completed. By shifting researchers' incentives towards undertaking more uncertain projects and/or allowing them to allocate more

[4]This mechanism is in line with Krieger et al. (2022) where, when firms receive cash windfalls, they invest in riskier drug candidates outside of their existing portfolio.

resources to a project, we posit that getting additional computing resources might increase the research space of researchers' projects.

Overall, an increase in the allocation of computing resources may lead researchers to pursue a different type of science, the nature of which we explore empirically.[5]

3 Setting

To investigate how constraints to supercomputing power impact innovation, we examine XSEDE, an NSF-funded initiative that provides a unique and valuable context for our study.

3.1 XSEDE

XSEDE was an initiative funded by the National Science Foundation (NSF) between 2009 and 2022 for a total of \$257M.[6] XSEDE was one of the most advanced, powerful, and robust collections of integrated digital resources and services in the world for open scientific research and was designed to provide researchers across various disciplines with access to advanced computing resources. XSEDE established a distributed network of high-performance computing resources located across the United States. These supercomputers enabled researchers to tackle computationally intensive scientific challenges that were previously beyond reach.

Table 1 shows the ten most used computing resources during our study period. Their associated total usage is expressed in the common metric used by XSEDE called Service Unit (SU), which can be roughly equated to "core-hours," and allows one to compare computing power across XSEDE resources that vary significantly in computational specifications.[7] Figure 1 shows the location of the resources listed in Table 1 in red, as well as the location of other resources used as part of the program in blue.

XSEDE played a pivotal role in facilitating groundbreaking discoveries across various scientific domains. One notable example of XSEDE's impact is its support for the Laser Interferometer Gravitational-Wave Observatory (LIGO) project. LIGO, a collaboration between Caltech and MIT, made history by detecting gravitational waves for the first time in 2015, confirming a key prediction of Einstein's theory of general

[5]Our discussion focuses on researchers' altering the projects they pursue when provided with additional resources. We cannot rule out the possibility that additional resources serve to reveal the pre-existing prioritization of projects without influencing choices. However, the core implications hold true even if one interprets the scenario as revealing an underlying project order.

[6]Following its success, XSEDE was replaced in 2022 by "ACCESS" (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support).

[7]XSEDE converts requests, recommendations, and allocations to SUs using resource-specific conversation factors that are based on the LINPACK benchmark, a commonly accepted measure that captures a resource's ability to quickly solve problems.

relativity. XSEDE provided essential computational resources and expertise to process and analyze the vast amounts of data generated by LIGO's detectors, contributing to this discovery. Additionally, XSEDE supported research in astrophysics, enabling researchers to conduct simulations of cosmic phenomena, such as the formation of galaxies and the behavior of black holes, leading to deeper insights into the structure and evolution of the universe. In climate science, XSEDE-supported research improved climate models and weather forecasting capabilities, aiding in disaster preparedness and climate change mitigation efforts. Furthermore, XSEDE played a crucial role in drug discovery and molecular simulations, providing researchers with the computational power needed to screen potential drug candidates, simulate molecular interactions, and accelerate the development of new therapeutics for various diseases.

3.2 Resource Allocation Process

Research allocations were available to researchers affiliated with U.S.-based organizations, with the vast majority of these researchers being associated with universities or other research institutions. Specifically, in our sample, 94.1% of researchers are associated with a university, 5.6% with a non-university research institution, and 0.3% with a commercial entity. Proposals, which contain computing allocation *requests*, were accepted quarterly. The lead researchers - hereafter referred to as Principal Investigators (PIs) - were required to include scientific background, research objectives, and a detailed resource usage plan covering code performance timings, resource usage details, scaling information, and benchmark runs on the requested resources.[8] Following a PI's initial allocation, they can request additional resources in following quarter(s) via a "renewal." Renewal proposals are evaluated in the same way but also ensure that the prior allocation(s) led to some progress. See Appendix C for more details on the allocation process.

Figure 2 shows an example of a proposal in the field of Astronomic Sciences, where the PI requested 476,372 SUs on the TACC machine "Stampede 2." Figure 3 shows the summary table of the detailed calculations made by the PI to justify his computing resources' request. The proposal also includes a description of the reason why XSEDE is critical for the PI's research agenda and emphasizes the limited availability of outside options. In particular, it highlights that the amount of computing resources that he needs is too high to be accommodated by his university. This echoes insights gathered from our interviews with other PIs who have used XSEDE for their research. One researcher, who runs simulations to understand the impact of pollution and heat dispersion, emphasized that XSEDE was critical for his work because typical university clusters are too small. Another researcher, focused on using genetic material to unravel evolutionary history, noted that

[8] Such benchmarking is typically done through a small "startup" allocation that allow researchers to run pilot code.

the computational demands of his work were so high that only XSEDE could provide the necessary resources, making it indispensable for his research.

For each of the quarterly windows of submission, the XSEDE Resource Allocation Committee (XRAC) is responsible for reviewing requests. On average, around 70% of projects are accepted and thus are awarded some computing power.[9] This committee includes about 60 computational researchers with expertise in various fields covering the diversity of fields of science encountered at XSEDE, though the most represented fields such as Material Research are more populated. Reviewers serve for a three-year term. The role of this team is to conduct an impartial and independent review of the resource allocation requests in order to make a *recommendation* regarding the amount of resources that the project (or "grant") would need to be completed. Importantly for our empirical design, reviewers are asked to gauge requests only on their own merit (i.e., what does the project require) and are specifically not informed of the overall amount of resources available nor the other requests across XSEDE's systems.[10]

After each meeting, the 'Allocations Team' meets to reconcile the recommendations made by the XRAC with the overall available supply of resources. This process starts with a "balancing" phase, where allocations from over-recommended resources are moved to architecturally equivalent under-subscribed resources until the excess demand is balanced equally across computing platforms. Next, this is followed by a "reconciliation" phase where recommendations are reduced through a formulaic solver in order to determine final *allocations*. The formula gives priority to NSF-funded projects, which represent about 50% of the proposals (see Appendix C.2 for more details on the reconciliation formula). Importantly, the only project-related characteristics considered in the reduction phase are (i) the recommendation made by the XRAC and (ii) the share of the project funding that is supported by the NSF.

Figure 4 summarizes the XSEDE allocation process. Figure 5 shows aggregate requests (i.e., demand), recommendations, and final allocations (i.e., recommendations after considering supply available) for each quarter of our study period.

4 Empirical Strategy

Empirically, we want to estimate the impact of the amount of computing resources allocated to researchers on scientific output (rate and characteristics). In the ideal experiment, we would assign each researcher (i.e.,

[9]Rejected projects are typically accepted in a future quarter, so there are very few outright rejections. This paper focuses on accepted proposals.
[10]Appendix C.1 provides excerpts from the *XSEDE Reviewer Manual*.

grant) a random number of computing hours that is weakly less than what it requires (i.e., randomly assign a level of constraint). We would then observe the subsequent output associated with the grant, which would allow us to estimate the effect of the marginal computing resource (i.e., marginal relaxing of constraint) on each researcher's productivity. While we cannot run such an experiment, we leverage the allocation process of XSEDE in order to get plausibly causal estimates.

In most similar settings one could study, the typical threat to a causal interpretation is that projects and researchers with a higher expected productivity will receive higher computing allocations, which could bias estimates. For instance, these projects would produce additional papers regardless of the quantity of resources they receive, biasing our quantity estimate upwards. We overcome this by exploiting the fact that, conditional on the recommendation from the XRAC and the percentage of project funding that is supported by the NSF, each grant's final allocation is a function of how busy the XSEDE's resources are, which is exogenous to the expected productivity of the grant. In other words, the amount of resources recommended to other projects - which influences the degree to which the XSEDE resources are constrained - is orthogonal to any feature of the focal grant once we control for the grant-related parameters included in the reconciliation formula: the amount of computing resources recommended by the XRAC and the share of the project funding supported by the NSF.

Our regressions of interest are of the form:

$$Y_i = \beta_0 + \beta_1 \ \%Allocated_i + \delta \ \ln(Recommended_i) + \gamma \ \%NSF\_Funding_i + \mathbf{X_i} + \varepsilon_i \tag{1}$$

where i indexes grants. Our treatment variable is $\%Allocated_i = \frac{Allocated_i}{Recommended_i}$ which captures the fraction of resources allocated to a grant compared to what it would have objectively needed, as assessed by the XRAC. This specification allows us to interpret the coefficient β_1 as the impact of receiving more or less computing resources relative to what a researcher would have needed.[11] Note that by definition of the XSEDE Allocation process, this ratio is less than or equal to one (i.e., all projects are weakly constrained). We control for the recommendation made by the XRAC with the logarithm of $Recommended_i$. One can think of this variable as capturing a combination of project quality and scale, where larger projects will inherently exhibit different output quantity and characteristics than smaller projects. The variable $\%NSF\_Funding_i$ accounts for the share of the project funding that is supported by the NSF, since this is part of the reconciliation formula. We include field of science fixed-effects, since some fields might apply for specific resources which are deferentially

[11] In Appendix B, we confirm all results are robust to using the level of allocation, $\ln(Allocated_i)$, as the treatment variable.

more or less busy and these fields might also have different research norms (e.g., more papers or different research characteristics). We also include grant start year fixed-effects in order to control for changes in overall supply over time as well as the potential for changing research output and characteristics over time. Our model includes the interaction of field of science and grant start year fixed-effects which control for field specific time trends. We use robust standard errors clustered at the grant start year $\times$ field level because the reconciliation process is performed across architecturally similar machines.[12] We test that other observable characteristics of researchers do not coincide with the level of constraint a project has. In Table A1, we show that the main independent variable of interest, $\%Allocated$, is uncorrelated with pre-determined covariates related to the PI: seniority, institution, number of papers published before XSEDE and number of citations received before XSEDE.

Given that the XRAC has information about the amount of computing resources requested by researchers, one might worry that recommendations are "anchored" on researchers' requests, leading some researchers to receive higher recommendations because they requested more resources than what their projects needed. In order for this to create endogeneity issues, asking for more should be correlated with both (i) the outcomes (i.e., PIs who ask for more should be either more or less productive than the average, or produce research with different characteristics) and (ii) the main explanatory variable (i.e., PI asking for more should have a higher or lower ratio than the average). While we cannot observe which PIs strategically asked for more than their true need, we create a proxy for the share of over-requested resources by measuring the 'percent over-request, $\frac{requested_i - recommended_i}{recommended_i}$, since larger gaps between the recommendation and request *may* suggest that the request was inflated by the researcher. We then examine the correlation between this variable and the ratio variable, including the controls. Table A2 shows no significant correlation. As previously mentioned, renewal proposals are additionally evaluated by confirming the project has made some progress (i.e., producing a paper) using their initial allocation(s) – which, in theory, could be problematic for our quantity analysis if one assumes having a prior paper is correlated with future expected productivity. Because larger recommendations are reduced more, the ratio we observe would be an underestimate of the true ratio, biasing our quantity results downward. See Appendix Section C.1 for further discussion.

[12] Fields of science at a given point in time often require particular types of computing resources that are architecturally similar, leading to a reconciliation process that may be correlated within field-year.

5 Data and Descriptive Statistics

5.1 Data

To study our question, we use data about allocation awards covering each quarter between 2015 and 2022.[13] For each proposal, the data contains information at the computing resource level about the initial request submitted by the PI, the recommended amount suggested by the XRAC and the final allocation that was awarded. We drop resources corresponding to "storage" and convert request, recommended, and allocated information associated with different resources into Service Unit using conversion factors provided by XSEDE. We also observe information about the proposal field of science, as well as PI first name, last name, and organization. While computing resources are awarded for a year, PIs can request renewals, which go through the same allocation process as their initial submissions. Since both the initial request and any subsequent renewals are tied to the same project, we treat them collectively as a single XSEDE grant. We therefore collapse information at the XSEDE grant level, keeping only grants for which we have complete information about their resources and for which we observe their first request year. Each row of our dataset contains information about the total amount of requested resources made by the PI, the corresponding total amount recommended by the XRAC and the total amount of allocated resources approved by the XSEDE allocation team. In theory, we could have kept our dataset at the request level instead of collapsing it at the PI ("grant") level. However, the scientific output that we observe can only be attributed to PIs and their XSEDE grants, and not to individual quarterly requests, making *grants* the relevant unit of observation. This corresponds to 1,063 unique grants associated with 1,063 unique PIs.

In order to match each XSEDE grant to its scientific research production, we use two sources of data. We first perform a fuzzy match of our population of PIs to the Dimensions AI database using first name, last name and institution name.[14] Once PIs are identified, we retrieve their papers and keep those that reference XSEDE in the acknowledgments section.[15] Dropping book chapters, this gives us 3,856 papers. We also obtained access to the XSEDE internal publication database, where PIs can report the publications associated with their grant. From this, we retrieve 922 additional papers. Note that a PI's decision to report to the internal XSEDE database is voluntary. This is not an issue for our empirical design, as long as the

[13]We are missing complete information about proposals for the years before 2015.

[14]See `https://www.dimensions.ai/`. In recent years Dimensions AI has been widely used in economics and management research, see e.g., Tian & Smith (2024), Fry & MacGarvie (2024), Myers et al. (2023), Myers & Tham (2023).

[15]From our discussions with PIs who were part of the XSEDE program, acknowledging XSEDE in the papers is the norm. We may be missing papers that attempt to acknowledge XSEDE without specifically mentioning a variation of "XSEDE" or "Extreme Science and Engineering Discovery Environment," but to be conservative we capture papers that we are confident originate from XSEDE computing allocations.

decision to voluntarily disclose publications is not correlated with our treatment variable. In Table A3, we show that there is no correlation between our treatment variable and papers identified through the use of the internal database.

We then retrieve information about each paper using Dimensions AI based on DOI number or title. Our empirical analysis relies heavily on "concepts" from the Dimensions database. The Dimensions AI database uses machine learning techniques to derive concepts in papers' abstracts and rank them based on their relevance on a scale from 0 (not relevant) to 1 (very relevant).[16] In all analyses, we only consider concepts with a score above 0.5. Appendix D provides more detail on Dimensions concepts and our use of them.

Among the initial 1,063 grants, we observe scientific output for 720 of them. Among the 343 unmatched grants, 112 come from PIs we were unable to match or to uniquely match with Dimensions. For the remaining 231 grants, 1 of them has papers which acknowledged XSEDE but could not be uniquely matched to a grant because their PIs were associated with several grants, and 230 grants have PIs who did not report any paper to the internal XSEDE database nor acknowledge XSEDE in one of their publications. Of the 720 grants with scientific output, we drop 9 grants whose ratio is strictly higher than 1 due to XSEDE data entry errors (see Table 2 for the successive sampling restrictions). Our final sample hence includes 711 unique grants. In order to assess whether matched grants are a selected sample, we construct an indicator equal to 1 if the grant is part of the 711 grants that constitute our main sample and 0 otherwise and we regress this measure on the main independent variable %*Allocated* and the main controls. Table A4 shows no statistically significant relationship.

5.2 Descriptive Statistics

Table 3 shows some descriptive statistics about our sample of 711 grants. On average, researchers request 61.5 million service units (SU). The XRAC makes lower recommendations on average, equal to 41.7 million SU and researchers gets a final average allocation of 35.7 million SU, with large variation in this number. On average, the share of allocated resources compared to the recommendation made by the review committee is 90%, but this variable is skewed. This can be seen in Figure 6 which shows the histogram of the treatment variable (*% Allocated*). On average, each grant is associated with 6.4 papers. Figure 7 shows the 15 most common fields of science in our sample. The most represented field is "Materials Research" with 56 grants, followed by Biophysics (48 grants) and Materials Engineering (41 grants).

[16] Several papers use MeSH terms to analyze the research topics of scientific papers (Azoulay et al. 2011, Myers 2020). MeSH terms are used for publications in PubMed which covers the biomedical literature. The use of Dimensions AI concepts allows us to extend this analysis to all the fields of science in our dataset. Appendix D shows examples of papers' concepts and abstracts.

6 Results

6.1 Quantity of Science

Our conceptual framework discusses the relationship between the amount of computing resources available to a researcher and the *quantity* and *type* of science they produce. In order to study the impact of additional resources on *quantity*, our first outcome of interest is the log of the total number of publications associated with each grant. Column (1) of Table 4 shows the results of regressions of the log of this outcome on the share of allocated resources received by grants.[17] Column (2) instead uses the inverse hyperbolic sine transformation. All columns control for the amount of computing resources recommended by the XRAC and the share of funding coming from the NSF. All specifications include grant start year fixed effects, field of science fixed effects and grant start year × field fixed effects. Both columns show a strong positive relationship. Using column (1) as our preferred specification, our results imply that a 10 percentage point increase in the percentage of allocated resources leads to approximately a 15% increase in the number of papers produced. Equivalently, an increase from the 10^{th} to the 75^{th} percentile in the share of allocated resources leads to a 60% increase in the number of papers produced.[18] This result implies that scientific output increases as resources constraints are relaxed.

6.2 Type of Science

In what follows, we explore the consequences of an increase in the share of allocated computing resources on the type of science that researchers pursue. We successively discuss whether such increases leads to: (i) researchers pursuing less traveled research avenues (ii) researchers studying topics beyond their expertise (iii) researchers broadening the scope of their research.

6.2.1 Less traveled research avenues

First, an increase in the share of computing resources could increase researchers' propensity to work on less-studied topics. For each paper, we calculate the number of times each concept it entails has been used by other PIs, excluding all papers from the focal PI, and average this measure at the paper level.[19] The higher this number, the more the paper entails concepts that have been used by other researchers too – these are

[17] Since our outcome does not contain zeroes (each grant in our sample is associated with at least one paper), we avoid issues described in Chen & Roth (2024).

[18] The 10th percentile of $\%Allocated$ equals 59% and the 75th percentile of $\%Allocated$ equals 1.

[19] We normalize this number to account for publication year following Perry & Reny (2016). This allows us to account for the fact that concepts used in more recent papers had less time to be used by other researchers.

"more studied" concepts. The lower this number, the more the paper entails concepts that have been less studied by other researchers – these are "less frequently-studied" concepts. We then collapse this to the grant level using different measures of interest.[20] First, we look at the average within grant, which gives us information about how well-studied the topics of the grant are on average. We also gain further insights about the overall distribution by considering the minimum within grant, i.e., the average use of concepts by other researchers for the least "well-studied" paper. This tells us about the effect on papers which entail the least-studied concepts. Finally, we also consider the maximum within grant, i.e., the average use of concepts by other researchers for the most "well-studied" paper. This gives us information about the effect on papers which entail the "most-studied" concepts.

Table 5 shows the results, with column (1) using the average as outcome, column (2) using the minimum as outcome and column (3) using the maximum as outcome. All columns control for the amount of resources recommended by the XRAC, as well as the share of funding coming from the NSF. All specifications include grant start year fixed effects, field fixed effects, and grant start year × field fixed effects. Column (1) displays a negative and statistically significant coefficient, showing that the higher the share of allocated resources, the less popular are the topics studied on average. The coefficient implies that a 10 percentage point increase in the share of allocated resources leads to the study of concepts that are about 5% less studied by other researchers. Such a relationship could be driven by a shift in the entire distribution of papers within a grant (i.e., every paper explores less studied concepts) or could be driven by changes at the tails (e.g., the papers that study the least (or most) popular topics explore more niche topics). Column (2) shows that this result is driven by papers in the left tail of the distribution: for the paper with the least studied topics, a 10 percentage point increase in allocation leads to the study of concepts that are about 8% less studied by other researchers. Column (3) shows no change at the right tail of the distribution. Said differently, an increase in the share of computing resources leads researchers to explore less-studied concepts, driven particularly by shifts in papers that focus on the least-studied topics.

This negative relationship could be driven by (i) the study of topics that were initially less popular but rising in popularity (ii) the study of topics that were popular but lose attractiveness or (iii) the study of topics that have a persistently lower level of use by others. In order to investigate this, we begin by calculating for each paper the number of times each concept it entails has been used by other PIs *prior to*, and separately *after*, the paper publication year and average this measure at the paper level. This allows us to study a relative trend in use by others before versus after publication year to disentangle (i)-(iii). We then reiterate

[20] We use the log of this measure, which is strictly positive.

our previous analysis, considering as outcome the mean value of this measure across all papers associated with a grant, the minimum value of this measure across all papers associated with a grant, and the maximum value of this measure across all papers associated with a grant. Results are presented in columns (4) to (6) using the '*use prior*' measure and show similar results as columns (1) to (3). Results presented in columns (7) to (9) use the '*use after*' measure and again show similar results.[21] This suggests that when researchers are allocated a higher share of resources, they study topics that are persistently less popular – i.e., less conventional.

We next study whether the concepts used are more or less recently introduced to the scientific literature as researchers get allocated additional computing resources – e.g., are these "less studied" concepts also "newer to the world?" We take each concept associated with a paper and calculate the first year we see this concept appearing among XSEDE researchers.[22] We then average this measure at the paper level. This gives us information about the average 'vintage' of the concepts used in a paper. The higher this measure, the more recent the concepts it includes are. The lower this measure, the older the concepts it studies are.

Column (1) of Table 6 uses as outcome the average of this value across all papers associated with a grant. This measure proxies how recent the concepts used by a grant are on average. Columns (2) and (3) use as outcome respectively the minimum and maximum of this value across all papers associated with a grant, which gives us information about the paper with the oldest and the most recent concepts, respectively. All specifications control for the amount of resources recommended by the XRAC, the share of funding coming from the NSF, grant start year fixed effects, field fixed effects, and grant start year $\times$ field fixed effects. Column (1) shows no significant effect, implying that an increase in the share of computing resources does not impact the *average* age of concepts associated with a grant. We also do not find any significant result in column (2), implying that an increase in the share of computing resources does not impact the age of concepts associated with the "oldest vintage" papers. However, column (3) shows a positive and statistically significant estimate at the 10% level, implying that an increase in the share of computing resources leads to the use of more recent concepts associated with the "newest vintage" papers.

This suggests that an increase in the share of computing resources leads researchers to advance their work into newer territories, as it's not the "oldest vintage" papers that change, but rather the "newest vintage" ones that explore even newer ideas. More specifically, column (3) displays that a 10 percentage point increase

[21] We conduct an t-test and confirm that these coefficients are statistically indistinguishable from those in columns (7) vs (4) and (9) vs (5), respectively.

[22] Since we do not observe the *universe* of scientific articles in our data, we cannot observe with certainty when the topic was initially studied overall. We use the first time we observe a topic in XSEDE as a good proxy for its appearance in the computational scientific literature. Further, we have no reason to suspect a correlation between when concepts appear among XSEDE researchers, relative to the universe, and a grant's computing allocation.

in %*Allocated* is associated with the study of topics that are about 4 years newer on average. Combining this results with Table 5 suggests that as resource constraints are relaxed, the concepts being studied are less common and newer to the world.

Taken together, we find strong evidence consistent with the fact that as resource constraints are relaxed, researchers choose to study topics that are less popular and are newer to the world.

6.2.2 Topics beyond expertise

An increase in the share of computing resources could also increase researchers' propensity to "pivot," i.e., to work on topics that they haven't studied previously. To analyze this, we calculate for each paper in our sample the share of concepts that have not been used by the focal PI in the years preceding their entry into XSEDE. We then study the average, minimum, and maximum of this measure at the grant level. Conceptually, we are capturing the degree of "pivoting" or "exploration" by each researcher, studying topics that they are less familiar with. Using the average allows us to capture the general tendency of researchers to explore new topics as computing resources increase. Using the minimum within a grant helps us see what happens for papers that study the most familiar topics. Using the maximum gives us information about papers that focus on the most unfamiliar topics, i.e., the most "exploratory" papers.

Table 7 shows the results. Columns (1)-(3) use the pool of papers written by the researcher in the two years preceding her entry into XSEDE to define the initial pool of papers with which we compare papers written as part of XSEDE, columns (4)-(6) looks at three years pre-XSEDE entry, and columns (7)-(9) look at five years pre-XSEDE entry. Columns (1), (4), and (7) look at the mean share within grant. Columns (2), (5), and (8) look at the minimum and columns (3), (6), and (9) look at the maximum. We find no significant effects on the mean nor minimum measures but strong positive results on the maximum value within grant. This implies that as resource constraints are relaxed, researchers explore 'new-to-them' topics among the most exploratory of their papers. This finding is consistent with the idea that the relaxing of computing constraints allows researchers to pursue more "exploratory" work that they do not seem to pursue in more constrained environments.

6.2.3 Research breadth

Finally, an increase in the share of computing resources could increase the research breadth of projects pursued by researchers. To explore this, we study the effect of an increase in the share of allocated computing resources on the number of unique concepts used by papers associated with a grant. We first calculate the number of unique concepts associated with each paper, which informs us about the diversity of ideas it embodies. Appendix D shows an example of this measure. Our first outcome averages this measure at the grant level. We use this as a proxy for the breadth of research associated with a grant: the higher the average number of concepts, the broader the research on average.[23] We also use as an outcome the minimum of this value across all papers associated with a grant. We use this as a proxy for the breadth of the "narrowest" papers associated with a grant. Similarly, we use as outcome the maximum of this value across all papers associated with a grant. We use this as a proxy for the breadth of the "broadest" papers associated with a grant.

Table 8 shows the results. All columns include grant start year fixed effects, field fixed effects, and grant start year × field fixed effects. Column (1) studies the average within grant and shows a significantly positive coefficient, implying that a 10 percentage point increase in the share of allocated resources leads to a 5% increase in the number of unique concepts associated with a grant. This indicates that a higher share of allocated resources leads researchers to expand their research breadth. Column (2) studies the minimum within grant and shows no effect. Column (3) uses the maximum within grant as the outcome and shows that the result in column (1) is driven by the "broadest" papers becoming broader. This implies that as resource constraints are relaxed, researchers work on broader papers and specifically expand the broadest ones.

6.3 Journal Quality and Citations

Thus far, we find evidence that as researchers receive more computing resources, they write more papers. We also find that researchers with additional resources study newer, less popular topics, study topics beyond their direct expertise and broaden the scope of their projects. These findings allow us to closely capture the impacts of computing resources on researcher strategies since they are products of the research undertaken by the researcher. In this section we explore downstream outcomes that the researchers have less control over. While we cannot directly tie any specific change in the type of science to journal placement and forward citations, we empirically explore how an increase in the share of computing resources impacts these outcomes. We use journal placement as our measure of *ex ante* quality and normalized forward citations as our measure of scientific impact.

First, we study the impact of additional resources on the quality of journal placement for papers associated with a grant. Specifically, we exploit information about the SNIP (Source Normalised Impact per Paper) value of the journal associated with each paper. SNIP measures the average citation impact of the publications of a

[23] In contrast to the journal breadth measure used in Azoulay et al. (2011), our measure more directly captures the researchers pursuit of broad topics separately from the impact in the scientific community.

journal and is therefore a good proxy of journal quality. Table 9, column (1) looks at the average SNIP value of papers associated with a grant. Columns (2) and (3) repeat this exercise by considering respectively the minimum and maximum SNIP value of papers associated with a grant. Overall, we do not find any significant estimate, suggesting that an increase in computing resources may not have a significant impact on the *ex ante* quality distribution. This suggests that an increase in the share of allocated resources does not lead to a decrease in inherent quality.

Next, we study the impact of an increase in the share of computing resources allocated to researchers on citations. In order to control for publication year and the field of science associated with each paper, we use an adjusted citation measure provided by Dimensions AI that calculates the scientific influence of a paper by dividing the number of citations it received by the average number of citations received by papers published in the same year and field of research. A value of one means that a paper receives exactly the average number of citations expected for its publication year and field. Since this measure already accounts for field differences, we do not control for field FE or field $\times$ year FE in these analyses.[24] This essentially allows us to compare the citations received by papers in our sample relative to the *universe* of papers, while normalizing by publication year and field. This measure alleviates potential concerns about truncation issues when studying papers published relatively recently, since papers are compared within field and publication year. We then collapse this measure at the grant level, calculating the mean, minimum, and maximum across papers. Results are presented in Table 9. Column (4) uses as outcome the average number of citations received by papers associated with each grant. Columns (5) and (6) repeat this exercise by using the number of citations associated with the least (respectively most) cited paper associated with a grant. Columns (4) and (6) show non-statistically significant estimates, while column (5) shows a strong negative relationship. This suggests that, while an increased in the share of computing resources does not lead to a significant change in average scientific impact or reception, it does lead to a significant decrease in the number of citations received by the least-cited papers associated with a grant.[25] Given the results in columns (1)-(3) and our findings about the study of newer topics, we are cautious to interpret this finding as representative of lower quality papers *per se*. While this is possible, it is also logical to interpret this finding as a consequence of the changing characteristics of the work produced which may correlate with less follow-on work by other researchers.

[24] We still control for grant start year FE since the Dimensions normalization is by paper publication year, which is different than grant start year.

[25] In unreported analyses, we construct an alternative field-year normalized citation measure that only uses our sample rather than the universe of papers and results are qualitatively similar.

7 Discussion

In this paper, we study how constraints on supercomputing resources, a key input to modern innovation, affects the quantity and type of innovation produced. At a high level, innovation is a critical driver of organizations' competitive advantage and overall economic growth (Romer 1990, Schumpeter 1934, Aghion et al. 2005) that results from the combination of various inputs (Henderson & Clark 1990, Fleming 2001). Understanding how constraints on key inputs, such as computing power, influence the rate and direction of innovation at the individual level is essential for optimizing resource allocation decisions, both in academic and industrial contexts. By examining the allocation of supercomputing resources among researchers, this paper provides some of the first empirical evidence into how variations in resource constraints can influence both the quantity and the characteristics of output of individuals.

We posit that as resource constraints are relaxed - i.e., as researchers receive more computing resources - output will increase and the type of science that researchers pursue may change. We then empirically investigate the relationship between resource constraints and research output in the context of XSEDE. By leveraging its specific allocation process, we are able to observe variations in the share of computing resources that researchers receive compared to what they would have needed, creating variations in how 'constrained' they are.

Our findings indicate that an increase in the share of allocated resources leads to an increase in research output, as evidenced by the increase in the number of papers published. We then explore changes to research direction. Our findings indicate that a higher share of allocated resources leads researchers to work on newer and less popular topics. We also find evidence that researchers shift toward topics that they themselves have not previously worked on. Finally, we find that researchers who receive more computing resources tend to broaden the scope of their research. We also find a decrease in citations but no significant change on journal quality, suggesting that these directional changes might lead to less follow-on work by other researchers. Importantly, these shifts are particularly salient for the *least* conventional papers, those that study the *newest* topics, the *most* exploratory papers, and the *broadest* papers, respectively – implying the relaxing of these constraints actually drives researchers to "spread their wings" and push boundaries.

Our findings suggest that computing resources, key constrained input to the innovation process, significantly impacts both the rate and direction of innovation. Our study expands our understanding of the role of resource constraints on the characteristics of innovation and highlights the multifaceted nature of the relationship between resource allocation, research strategies, and characteristics of the output produced. In particular, our findings underscore the importance of considering the nuanced impacts of resource allocation strategies on research outcomes and the need for continued exploration of the underlying mechanisms driving these dynamics.

Future work could delve deeper into the relationship between researcher characteristics and the impact of constraints or their heterogeneous value of computing power. Additionally, exploring how other types of resource constraints impact the characteristics of innovative output could provide a more comprehensive understanding of how our findings apply to other constrained inputs. Finally, future work should consider how access to complementary resources, such as data, interacts with the effects we highlight.

Practically, this work has clear policy and managerial implications. Our study highlights the need for R&D managers and government programs with limited resources to carefully consider how resource allocation decisions affect the characteristics of research output. For R&D managers, particularly those in firms where innovation is critical to competitive advantage, our study provides valuable insights into how resource allocation decisions can influence the direction of research. Managers should be aware that providing less computational resources can lead to shifts in research focus, with researchers engaging in different types of projects that are less novel and narrower. The implications of our findings are also particularly relevant for policymakers who are involved in the allocation of resources to scientists. Our study underscores the importance of recognizing the impact of computing resource allocations on the type of science ultimately produced. Since firms also often build on academic research as the foundation for new technologies, the type of science produced in academia directly influences the innovations that firms can develop (Marx & Fuegi 2020, Arora et al. 2018, Aghion et al. 2008, Cohen et al. 2002). This makes the quality and direction of scientific output a crucial factor in shaping future technological advancements.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. & Howitt, P. (2005), 'Competition and innovation: An inverted-u relationship', *The quarterly journal of economics* **120**(2), 701–728.

Aghion, P., Dewatripont, M. & Stein, J. C. (2008), 'Academic freedom, private-sector focus, and the process of innovation', *The RAND Journal of Economics* **39**(3), 617–635.

Agrawal, A., Catalini, C., Goldfarb, A. & Luo, H. (2018), 'Slack time and innovation', *Organization Science* **29**(6), 1056–1073.

Arora, A., Belenzon, S. & Patacconi, A. (2018), 'The decline of science in corporate r&d', *Strategic Management Journal* **39**(1), 3–32.

Arora, A. & Gambardella, A. (2005), 'The impact of nsf support for basic research in economics', *Annales d'Economie et de Statistique* pp. 91–117.

Arrow, K. (1962), Economic Welfare and the Allocation of Resources for Invention, Princeton University Press, Princeton, pp. 609–626.

Atanassov, J. (2016), 'Arm's length financing and innovation: Evidence from publicly traded firms', *Management Science* **62**(1), 128–155.

Azoulay, P., Ding, W. & Stuart, T. (2007), 'The determinants of faculty patenting behavior: Demographics or opportunities?', *Journal of economic behavior & organization* **63**(4), 599–623.

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

Azoulay, P., Graff Zivin, J. S. & Manso, G. (2011), 'Incentives and creativity: evidence from the academic life sciences', *The RAND Journal of Economics* **42**(3), 527–554.

Babina, T., He, A. X., Howell, S. T., Perlman, E. R. & Staudt, J. (2023), 'Cutting the innovation engine: how federal funding shocks affect university patenting, entrepreneurship, and publications', *The Quarterly Journal of Economics* **138**(2), 895–954.

Barney, J. (1991), 'Firm resources and sustained competitive advantage', *Journal of management* **17**(1), 99–120.

Berrone, P., Fosfuri, A., Gelabert, L. & Gomez-Mejia, L. R. (2013), 'Necessity as the mother of 'green'inventions: Institutional pressures and environmental innovations', *Strategic Management Journal* **34**(8), 891–909.

Blomfield, M. & Vakili, K. (2023), 'Incentivizing effort allocation through resource allocation: Evidence from scientists' response to changes in funding policy', *Organization Science* **34**(1), 100–128.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R. & Riedl, C. (2016), 'Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science', *Management science* **62**(10), 2765–2783.

Boyabatlı, O., Leng, T. & Toktay, L. B. (2016), 'The impact of budget constraints on flexible vs. dedicated technology choice', *Management Science* **62**(1), 225–244.

Bronzini, R. & Piselli, P. (2016), 'The impact of r&d subsidies on firm innovation', *Research Policy* **45**(2), 442–457.

Catalini, C. (2018), 'Microgeography and the direction of inventive activity', *Management Science* **64**(9), 4348–4364.

Cerqueiro, G., Hegde, D., Penas, M. F. & Seamans, R. C. (2017), 'Debtor rights, credit supply, and innovation', *Management Science* **63**(10), 3311–3327.

Chen, J. & Roth, J. (2024), 'Logs with zeros? some problems and solutions', *The Quarterly Journal of Economics* **139**(2), 891–936.

Choi, B., Kumar, M. S. & Zambuto, F. (2016), 'Capital structure and innovation trajectory: The role of debt in balancing exploration and exploitation', *Organization Science* **27**(5), 1183–1201.

Cohen, W. M., Nelson, R. R. & Walsh, J. P. (2002), 'Links and impacts: the influence of public research on industrial r&d', *Management science* **48**(1), 1–23.

Conti, A. (2018), 'Entrepreneurial finance and the effects of restrictions on government r&d subsidies', *Organization Science* **29**(1), 134–153.

Conti, A., Dass, N., Di Lorenzo, F. & Graham, S. J. (2019), 'Venture capital investment strategies under financing constraints: Evidence from the 2008 financial crisis', *Research Policy* **48**(3), 799–812.

Danneels, E. & Sethi, R. (2011), 'New product exploration under environmental turbulence', *Organization Science* **22**(4), 1026–1039.

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

Ding, W. W., Levin, S. G., Stephan, P. E. & Winkler, A. E. (2010), 'The impact of information technology on academic scientists' productivity and collaboration patterns', *Management Science* **56**(9), 1439–1461.

Dongarra, J., Beckman, P., Moore, T., Aerts, P., Aloisio, G., Andre, J.-C., Barkai, D., Berthou, J.-Y., Boku, T., Braunschweig, B. et al. (2011), 'The international exascale software project roadmap', *The international journal of high performance computing applications* **25**(1), 3–60.

Dutt, N. & Lawrence, M. (2022), 'Learning to manage breadth: experience as repetition and adaptation', *Organization Science* **33**(4), 1300–1318.

Ferguson, J.-P. & Carnabuci, G. (2017), 'Risky recombinations: Institutional gatekeeping in the innovation process', *Organization Science* **28**(1), 133–151.

Fini, R., Perkmann, M. & Ross, J.-M. (2022), 'Attention to exploration: The effect of academic entrepreneurship on the production of scientific knowledge', *Organization Science* **33**(2), 688–715.

Fitzgerald, T., Balsmeier, B., Fleming, L. & Manso, G. (2021), 'Innovation search strategy and predictable returns', *Management science* **67**(2), 1109–1137.

Fleming, L. (2001), 'Recombinant uncertainty in technological search', *Management science* **47**(1), 117–132.

Foster, J. G., Rzhetsky, A. & Evans, J. A. (2015), 'Tradition and innovation in scientists' research strategies', *American Sociological Review* **80**, 875–908.

Franzoni, C. & Stephan, P. (2023), 'Uncertainty and risk-taking in science: Meaning, measurement and management in peer review of research proposals', *Research Policy* **52**(3), 104706.

Fry, C. & MacGarvie, M. (2024), 'Author country of origin and attention on open science platforms: Evidence from covid-19 preprints', *Management Science* **70**(8), 5426–5444.

Furman, J. L. & Teodoridis, F. (2020), 'Automation, research technology, and researchers' trajectories: Evidence from computer science and electrical engineering', *Organization Science* **31**(2), 330–354.

Glennon, B. (2023), 'How do restrictions on high-skilled immigration affect offshoring? evidence from the h-1b program', *Management Science* .

Hall, B. & Van Reenen, J. (2000), 'How effective are fiscal incentives for r&d? a review of the evidence', *Research policy* **29**(4-5), 449–469.

Henderson, R. M. & Clark, K. B. (1990), 'Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms', *Administrative Science Quarterly* **35**(1), 9–30.

Hill, R., Yin, Y., Stein, C., Wang, D. & Jones, B. (2021), 'Adaptability and the pivot penalty in science', *Working Paper* .

Hoegl, M., Gibbert, M. & Mazursky, D. (2008), 'Financial constraints in innovation projects: When is less more?', *Research Policy* **37**(8), 1382–1391.

Howell, S. T. (2017), 'Financing innovation: Evidence from r&d grants', *American economic review* **107**(4), 1136–1164.

Jacob, B. A. & Lefgren, L. (2011), 'The impact of research grant funding on scientific productivity', *Journal of public economics* **95**(9-10), 1168–1177.

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

Kim, S. (2023), 'Shortcuts to Innovation: The Use of Analogies in Knowledge Production', *Working Paper* .

Kong, D., Lin, C., Wei, L. & Zhang, J. (2022), 'Information accessibility and corporate innovation', *Management Science* **68**(11), 7837–7860.

Krieger, J. L., Myers, K. R. & Stern, A. D. (2023), 'How important is editorial gatekeeping? evidence from top biomedical journals', *Review of Economics and Statistics* pp. 1–33.

Krieger, J., Li, D. & Papanikolaou, D. (2022), 'Missing novelty in drug development', *The Review of Financial Studies* **35**(2), 636–679.

Lohr, S. (2019), At tech's leading edge, worry about a concentration of power, Technical report.

Luger, J., Raisch, S. & Schimmer, M. (2018), 'Dynamic balancing of exploration and exploitation: The contingent benefits of ambidexterity', *Organization Science* **29**(3), 449–470.

Mandler, M. (2017), 'The benefits of risky science', *The Economic Journal* **127**(603), 1495–1526.

Manso, G. (2011), 'Motivating innovation', *The journal of finance* **66**(5), 1823–1860.

Marino, A., Aversa, P., Mesquita, L. & Anand, J. (2015), 'Driving performance via exploration in changing environments: Evidence from formula one racing', *Organization Science* **26**(4), 1079–1100.

Marx, M. & Fuegi, A. (2020), 'Reliance on science: Worldwide front-page patent citations to scientific articles', *Strategic Management Journal* **41**(9), 1572–1594.

Mezzanotti, F. & Simcoe, T. (2023), Research and/or development? financial frictions and innovation investment, Technical report, National Bureau of Economic Research.

Millett, L. I. & Fuller, S. H. (2011), *The future of computing performance: game over or next level?*, National Academies Press.

Myers, K. (2020), 'The Elasticity of Science', *American Economic Journal: Applied Economics* **12**(4), 103–134.

Myers, K. & Lanahan, L. (2022), 'Estimating spillovers from publicly funded R&D: Evidence from the US Department of Energy', *American Economic Review* **Forthcoming**.
URL: *www.subsurfaceinsights.com*

Myers, K. R., Tham, W. Y., Thursby, J., Thursby, M., Cohodes, N., Lakhani, K., Mural, R. & Xu, Y. (2023), 'New facts and data about professors and their research', *arXiv preprint arXiv:2312.01442* .

Myers, K. & Tham, W. Y. (2023), 'Money, time, and grant design', *Working Paper* .
URL: $https://www.kylemyers.org/_files/ugd/73b908_16ea1f7955fd472683781172c6c8bb0f.pdf$

Nagaraj, A. & Tranchero, M. (2023), How does data access shape science? evidence from the impact of us census's research data centers on economics research, Technical report, National Bureau of Economic Research.

Nagle, F. & Teodoridis, F. (2020), 'Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration', *Strategic management journal* **41**(1), 55–85.

Nanda, R. & Rhodes-Kropf, M. (2017), 'Financing risk and innovation', *Management Science* **63**(4), 901–918.

Nelson, R. R. (1959), 'The economics of invention: A survey of the literature', *The Journal of Business* **32**(2), 101–127.

Payne, A. A., Siow, A. et al. (1999), *Does federal research funding increase university research output?*, Citeseer.

Perry, M. & Reny, P. J. (2016), 'How to count citations if you must', *American Economic Review* **106**(9), 2722–2741.

Romer, P. (1990), 'Endogenous Technological Change', *Journal of Political Economy* **98**(5, part 2), S71–S102.

Schumpeter, J. A. (1934), 'The theory of economic development'.

Sidhu, J. S., Commandeur, H. R. & Volberda, H. W. (2007), 'The multifaceted nature of exploration and exploitation: Value of supply, demand, and spatial search for innovation', *Organization Science* **18**(1), 20–38.

Singh, J. & Marx, M. (2013), 'Geographic constraints on knowledge spillovers: Political borders vs. spatial proximity', *Management Science* **59**(9), 2056–2078.

Stephan, P. E. (1996), 'The economics of science', *Journal of Economic literature* **34**(3), 1199–1235.

Thompson, N. C., Ge, S. & Manso, G. F. (2022), 'The importance of (exponentially more) computing power'.

Tian, T. Y. & Smith, E. B. (2024), 'Stretched thin: How a misalignment between allocation and valuation underlies the paradox of diversity achievement in higher education', *Administrative Science Quarterly* p. 00018392241247744.

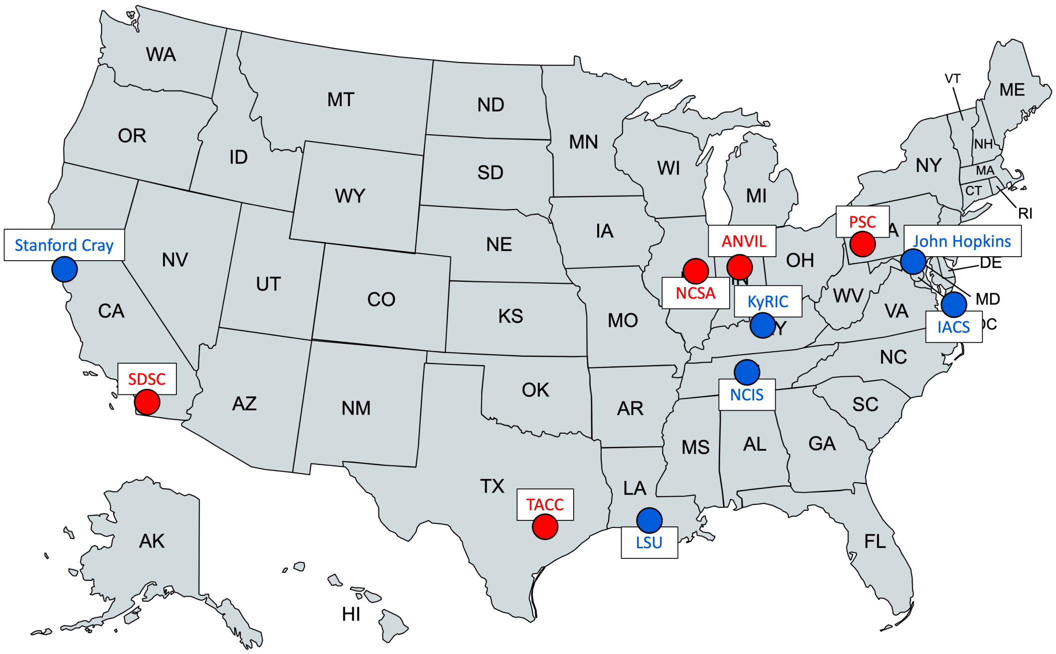
Tranchero, M. (2023), 'Finding Diamonds in the Rough: Data-Driven Opportunities and Pharmaceutical Innovation', *Working Paper* .

Zhou, K. Z. & Li, C. B. (2012), 'How knowledge affects radical innovation: Knowledge base, market knowledge acquisition, and internal knowledge sharing', *Strategic management journal* **33**(9), 1090–1102.

Zhuo, R. (2022), 'Exploit or explore? an empirical study of resource allocation in scientific labs', *Working Paper* .

Figures

Figure 1: Location of XSEDE resources



Notes: This table shows the resources that are associated with XSEDE. The ten most used resources, listed in Table 1 are in red and others are in blue.

Figure 2: Abstract and Scientific Background

Numerical Simulations of Protoplanetary Disks

Summary

We request a total of **476, 372 SUs** on the TACC machine STAMPEDE 2 SKX nodes to run and analyze numerical simulations of magnetohydrodynamic turbulence in the planet forming region of protoplanetary disks. We will use the well-tested ATHENA code to simulate a series of local, co-rotating radial regions of a model disk. To store the large datasets that will result from these calculations, we request $\mathbf{1.744 \times 10^4}$ **GB** of archival storage on the TACC RANCH system.

Notes: Yellow highlights added for emphasis.

Figure 3: Requested Computing Resources

Table 1: Proposed Simulations

R (AU)	β_z	O_B	# zones	# cores	$\frac{\text{zones}}{\text{core}}$	N_t	Days	**SUs**
5	10^3	1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
5	10^3	-1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
5	10^4	1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
5	10^4	-1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
5	10^5	1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
5	10^5	-1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
1	10^4	1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
10	10^4	1	256×512^2	4,096	16×32^2	1.59×10^7	33.5	$\mathbf{69,147}$
							Total:	553,172 SUs
							Other resources:	-76,800 SUs
							Final Total:	**476,372 SUs**

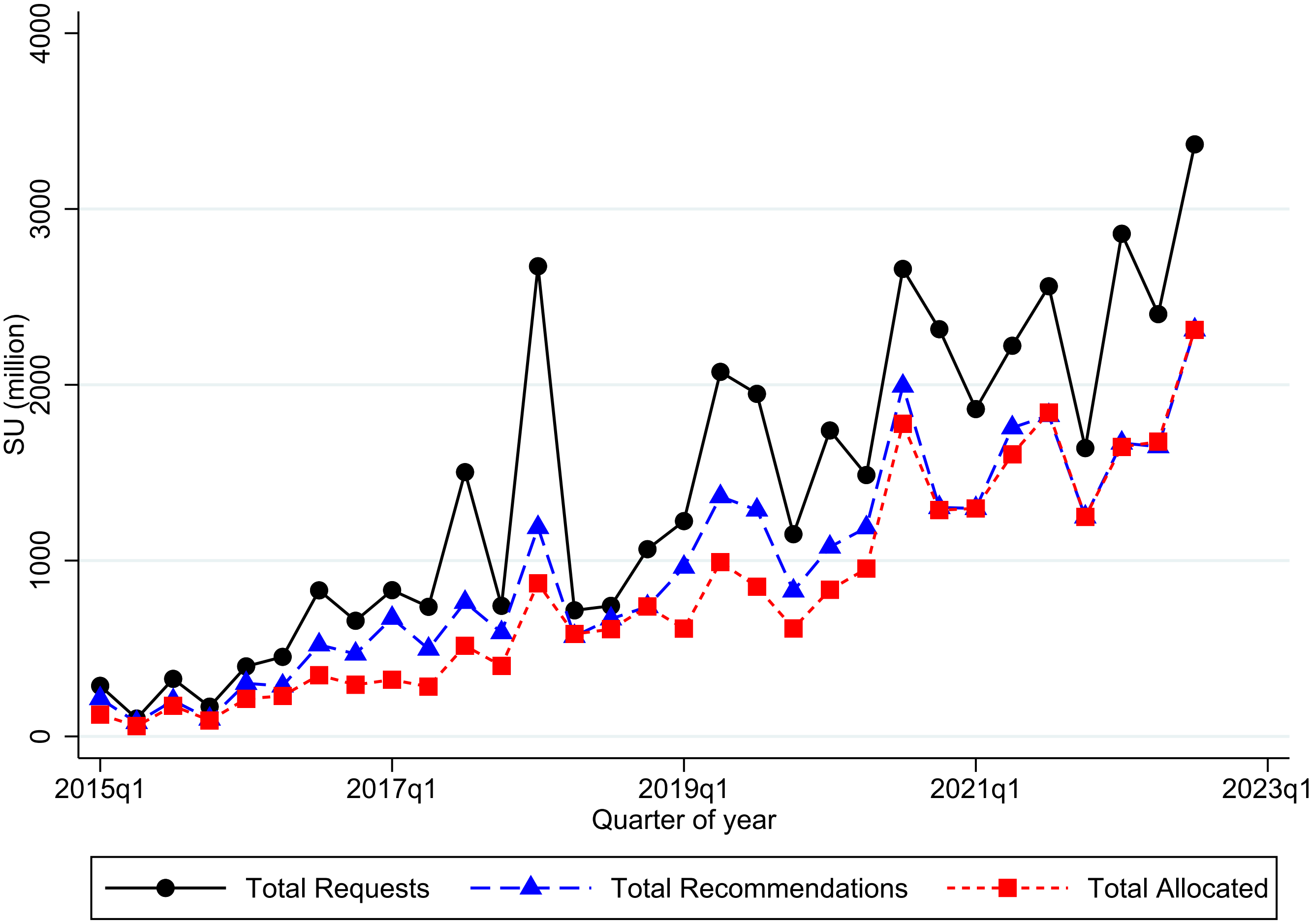
5. Computational Resources

Access to the TACC system STAMPEDE 2 is *essential* for the success of our research project. Locally, our group does have access to a moderate sized high performance computing system, SUMMIT. This machine has 9,120 general computing CPUs (plus GPU capability on other nodes) and is operated through the University of Colorado, Boulder. However, using this machine for the simulations proposed here is problematic for two reasons. First, the number of cores required for a single shearing box simulation is a substantial fraction of the maximum number of cores on SUMMIT. Thus, we could never expect this system to permit regular access to the 4,096 cores required for our calculations. Second, in consulting with the Research Computing division at the University of Colorado, it has become clear that we can acquire at most one million CPU-hours on this machine per year *per research group*, significantly less than what we require to carry out our proposed simulations.

Notes: Yellow highlights added for emphasis.

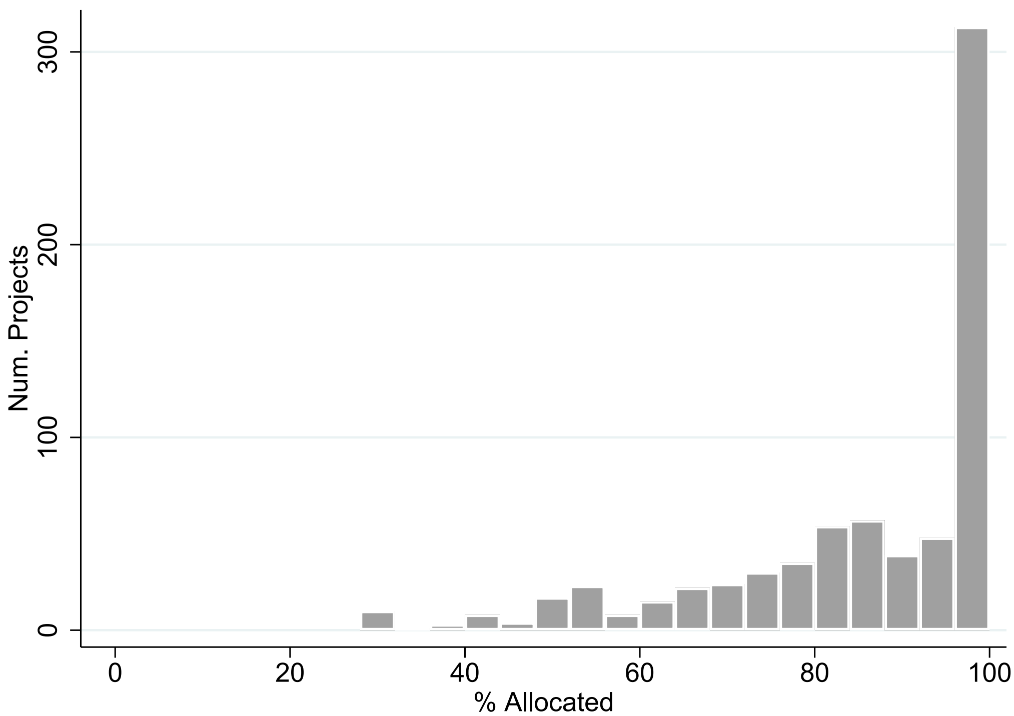
Figure 4: XSEDE Resource Allocation Process

Figure 5: Total Requests, Recommendation and Supply Over Time



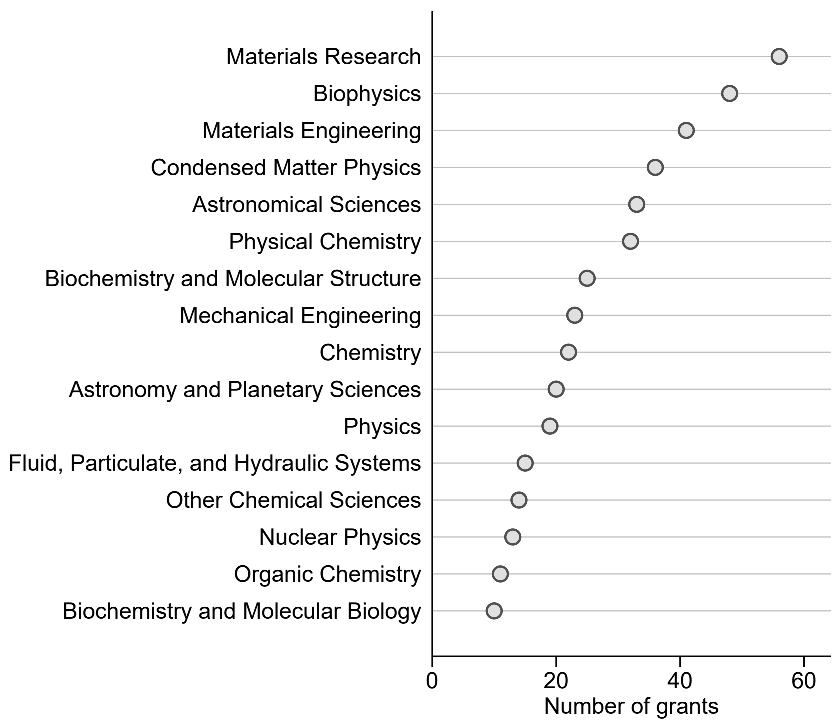
Notes: The x-axis shows each quarter between 2015 and 2022. 'Total Requests' represents the total amount of requested resources, 'Total Recommendations ' represents the total amount of recommendations made by the XRAC committee, and 'Total Allocated ' represents the total amount of allocated resources after reconciliation. Several resources (Indiana Jetstream, Kentucky Research Informatics, PSC Bridges-2, Purdue Anvil) were launched in 2021, leaving the XSEDE supply less constrained and explaining the near-overlap between the 'Total Recommendations' and the 'Total Allocated' in the latest quarters.

Figure 6: Histogram of the share of resources allocated



Notes: This figure reports the distribution of the treatment variable %Allocated, which represents the ratio of computing resources allocated to a grant compared to the amount of resources recommended by the XRAC.

Figure 7: Most common fields of science (top 15)



Notes: This figure reports the number of grants per field of science for the most common 15 fields of science.

Tables

Table 1: Ten Most Used Resources

Resource Name	SU (Millions)
TACC Dell/Intel Knights Landing (Stampede2)	27,454
SDSC Dell Cluster (Comet)	10,327
SDSC Expanse CPU	9,306
PSC Bridges-2 (Bridges-2)	8,493
TACC Dell PowerEdge (Stampede)	6,320
PSC Regular Memory (Bridges)	6,079
Purdue Anvil CPU	3,674
NCSA Delta GPU (Delta GPU)	2,340
IU/TACC (Jetstream)	2,264
PSC Bridges-2 GPU (Bridges-2 GPU)	1,965

Notes: This table shows the 10 most used resources over the period 2015-2022. Computing resources are all reported in Service Units (SU) which allows comparison across resources.

Table 2: Sample construction

Sample	Observations
Grants	1,063
Grants with identifiable PIs	951
Grants with a 1:1 correspondence with their PIs	950
Grants with at least one associated paper	720
Grants with a ratio$\leq$1	711

Notes: This table shows the successive steps that lead to the construction of the final sample with 711 observations.

Table 3: Summary statistics

	Mean	SD	Min	Max
Requested (Mil. SU)	61.5	144.2	0.0	1627.6
Recommended (Mil. SU)	41.7	91.1	0.1	930.7
Approved (Mil. SU)	35.7	80.1	0.1	862.9
% Allocated	0.9	0.2	0.3	1.0
Grant start year	2017.9	2.0	2015.0	2022.0
R1 universities	0.8	0.4	0.0	1.0
Junior PI	0.2	0.4	0.0	1.0
Number of papers per grant	6.4	9.3	1.0	97.0
Log(use)	1.0	0.8	0.0	11.8
Mean concept year	2004.4	4.7	1980.0	2023.0
Share of new concepts (2yr)	0.8	0.2	0.0	1.0
Average number of unique concepts	24.8	9.7	1.0	66.7
Observations	711			

Notes: All summary statistics are at the grant level. Log (use) is defined as the average number of times concepts used in papers associated with a grant are used by other researchers and is described in greater detail in section 6.2.1. Mean concept year is defined as the average year in which concepts used in papers associated with a grant are first observed and is described in greater detail in section 6.2.1. Share of new concepts (2yr) is constructed by calculating the share of concepts that have not been used by the focal PI in the two years preceding their first entry into XSEDE and is described in greater detail in section 6.2.2. *Average number of unique concepts* is the average number of unique concepts associated with a grant and is described in greater detail in section 6.2.3.

Table 4: Quantity of papers

	ln(papers)	asinh(papers)
	(1)	(2)
$\%Allocated_i$	0.902***	0.834***
	(0.258)	(0.241)
$\ln(Recommended_i)$	0.280***	0.264***
	(0.037)	(0.035)
$\%\text{NSF}\, Funding_i$	-0.070	-0.066
	(0.104)	(0.097)
Year FE	Yes	Yes
Field FE	Yes	Yes
Field × Year FE	Yes	Yes
Dep Var Mean	1.35	2.10
R-Sq	0.45	0.45
Observations	523	523

Notes: The unit of observation is a grant. In column (1), the outcome is the (log) number of papers associated with a grant. In column (2), the outcome is the inverse hyperbolic sine of the number of papers associated with a grant. The observation count shown is the count with remaining variation after fixed effects are included. %*Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. $\%NSF\, Funding$ is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

Table 5: Use of concepts by others

	Log(Use)			Log(Use Prior)			Log(Use After)		
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(7) Mean	(8) Min	(9) Max
$\%Allocated_i$	-0.677**	-1.780***	-0.373	-0.643**	-1.639***	-0.309	-0.780***	-2.281***	-0.355
	(0.279)	(0.471)	(0.313)	(0.290)	(0.464)	(0.323)	(0.279)	(0.475)	(0.295)
$\ln(Recommended_i)$	0.053	-0.134**	0.151***	0.033	-0.121**	0.129***	0.052	-0.034	0.168***
	(0.048)	(0.060)	(0.051)	(0.044)	(0.054)	(0.047)	(0.051)	(0.072)	(0.051)
%NSF $Funding_i$	-0.022	-0.087	-0.051	0.050	-0.018	0.024	-0.055	0.026	-0.100
	(0.118)	(0.146)	(0.136)	(0.101)	(0.145)	(0.121)	(0.132)	(0.146)	(0.147)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	-0.41	-1.27	0.07	-0.38	-1.20	0.10	-0.50	-0.99	0.05
R-Sq	0.67	0.55	0.63	0.68	0.57	0.64	0.69	0.63	0.65
Observations	465	397	465	462	383	462	448	309	448

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. In columns (1) to (3), we first calculate the average number of times concepts used in a paper are used by other PIs. Column (1) takes the mean of this value at the grant level, logged. Column (2) takes the min of this value at the grant level, logged. Column (3) takes the max of this value at the grant level, logged. Columns (4) to (6) repeat this exercise using the average number of times concepts used in a paper are used by other PIs prior to her entry into XSEDE. Columns (7) to (9) repeat this exercise using the average number of times concepts used in a paper are used by other PIs after her entry into XSEDE. %*Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. %*NSF Funding* is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

Table 6: Concept Vintage

	Concept Year		
	(1) Mean	(2) Min	(3) Max
$\%Allocated_i$	1.259	-0.684	4.177*
	(1.492)	(2.118)	(2.164)
$\ln(Recommended_i)$	-0.359**	-1.250***	0.425**
	(0.179)	(0.249)	(0.215)
$\%\text{NSF}\,Funding_i$	-0.615	-0.639	-1.008
	(0.470)	(0.683)	(0.675)
Year FE	Yes	Yes	Yes
Field FE	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes
Dep Var Mean	2004	2000	2008
R-Sq	0.44	0.43	0.39
Observations	485	485	485

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. For each paper associated with a grant, we calculate the average year we observe the concepts it entails for the first time. Column (1) averages this measure at the grant level. Column (2) takes the min of this measure at the grant level. Column (3) takes the max of this measure at the grant level. $\%Allocated$ is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. $\ln(Recommended)$ is the log of the number of SUs recommended by XRAC. $\%NSF\,Funding$ is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

Table 7: Share of new concepts

	Vs prev. 2y			Vs prev. 3y			Vs prev. 5y		
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(7) Mean	(8) Min	(9) Max
$\%Allocated_i$	0.065	-0.070	0.119**	0.091	-0.025	0.143***	0.083	-0.036	0.141***
	(0.057)	(0.097)	(0.051)	(0.058)	(0.105)	(0.048)	(0.061)	(0.104)	(0.052)
$\ln(Recommended_i)$	-0.002	-0.044***	0.014**	-0.001	-0.042***	0.016**	-0.004	-0.046***	0.016**
	(0.008)	(0.015)	(0.007)	(0.008)	(0.014)	(0.007)	(0.008)	(0.014)	(0.007)
$\%\text{NSF}\ Funding_i$	0.006	0.021	-0.006	-0.001	0.011	-0.016	-0.001	0.013	-0.017
	(0.026)	(0.044)	(0.023)	(0.025)	(0.041)	(0.023)	(0.024)	(0.040)	(0.022)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	0.78	0.63	0.88	0.76	0.61	0.87	0.74	0.59	0.86
R-Sq	0.45	0.37	0.42	0.45	0.37	0.43	0.45	0.37	0.44
Observations	430	430	430	438	438	438	446	446	446

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. For each paper, we calculate the share of 'new-to-the-researcher' concepts. Columns (1) to (3) compare XSEDE papers with papers written in the 2 years that precede XSEDE entry. Columns (4) to (6) compare XSEDE papers with papers written in the 3 years that precede XSEDE entry. Columns (7) to (9) compare XSEDE papers with papers written in the 5 years that precede XSEDE entry. Columns (1), (4) and (7) take the average of this measure at the grant level. Columns (2), (5) and (8) take the min of this measure at the grant level. Columns (3), (6) and (8) take the max of this measure at the grant level. %*Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. $\%NSF\ Funding$ is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

Table 8: Research Breadth

	Log(Unique Concepts)		
	(1) Mean	(2) Min	(3) Max
$\%Allocated_i$	0.420**	-0.141	0.651***
	(0.168)	(0.337)	(0.174)
$\ln(Recommended_i)$	-0.009	-0.256***	0.045*
	(0.023)	(0.047)	(0.024)
$\%\text{NSF}\,Funding_i$	0.037	0.150	0.015
	(0.046)	(0.125)	(0.053)
Year FE	Yes	Yes	Yes
Field FE	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes
Dep Var Mean	3.09	2.32	3.37
R-Sq	0.35	0.42	0.35
Observations	475	475	475

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. We calculate for each paper the number of concepts it entails. Column (1) averages this measure at the grant level, logged. Column (2) takes the min of this measure at the grant level, logged. Column (3) takes the max of this measure at the grant level, logged. $\%Allocated$ is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. $\ln(Recommended)$ is the log of the number of SUs recommended by XRAC. $\%NSF\,Funding$ is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

Table 9: Journal quality and Number of normalized citations

	Log(SNIP)			Log(Citations)		
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max
$\%Allocated_i$	-0.054	-0.094	-0.018	-0.248	-0.610***	0.105
	(0.079)	(0.065)	(0.121)	(0.219)	(0.216)	(0.269)
$\ln(Recommended_i)$	0.018**	-0.010	0.059***	0.108***	-0.019	0.208***
	(0.009)	(0.008)	(0.014)	(0.022)	(0.025)	(0.027)
$\%\text{NSF}\ Funding_i$	-0.046*	-0.033	-0.062*	-0.011	0.004	-0.006
	(0.026)	(0.024)	(0.036)	(0.084)	(0.084)	(0.097)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes			
Field × Year FE	Yes	Yes	Yes			
Dep Var Mean	0.66	0.51	0.85	1.64	0.93	2.10
R-Sq	0.29	0.32	0.33	0.05	0.03	0.10
Observations	502	502	502	619	619	619

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. In column (1), the outcome is the (log) average SNIP value of papers associated with a grant. In column (2), the outcome is the (log) minimum SNIP value of papers associated with a grant. In column (3), the outcome is the (log) maximum SNIP value of papers associated with a grant. SNIP values are normalized to account for differences in citation norms over years. In column (4), the outcome is the (log) average number of citations associated with a grant. In column (5), the outcome is the (log) number of citations received by the least cited paper associated with a grant. In column (6), the outcome is the (log) number of citations received by the most cited paper associated with a grant. Citations associated with each paper are normalized to account for differences across fields and publication year. $\%Allocated$ is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. $\ln(Recommended)$ is the log of the number of SUs recommended by XRAC. $\%NSF\ Funding$ is the share of funding coming from the NSF. All columns control for the amount recommended. Columns (1) to (3) control for grant start year fixed effects, grant field fixed effects and grant start year × grant field fixed effects. Columns (4) to (6) control for grant start year fixed effects. Standard errors are clustered at the grant start year × field level.

A Tables: Robustness of empirical strategy

Table A1: Balance test

	%Allocated				
	(1)	(2)	(3)	(4)	(5)
Junior PI	-0.001				-0.009
	(0.017)				(0.017)
R1 Univ		-0.019			-0.006
		(0.019)			(0.020)
Nb papers pre-XSEDE			-0.000		-0.000
			(0.000)		(0.000)
Nb cites pre-XSEDE				-0.000	-0.000
				(0.000)	(0.000)
%NSF $Funding_i$	0.102***	0.098***	0.102***	0.102***	0.102***
	(0.018)	(0.017)	(0.019)	(0.019)	(0.019)
ln($Recommended_i$)	-0.025***	-0.027***	-0.024***	-0.024***	-0.024***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Field FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	0.85	0.85	0.85	0.85	0.85
R-Sq	0.51	0.49	0.52	0.52	0.52
Observations	501	523	494	494	494

Notes: The unit of observation is a researcher. *%Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. *Junior* is an indicator equal to 1 the PI has less than 8 years of experience. *R*1 is an indicator equal to 1 if the PI is affiliated with an R1 institution. *Nb papers pre-XSEDE* is the number of publications published by the PI before entering XSEDE. *Nb cites pre-XSEDE* is the number of citations received by the PI for publications published before entering XSEDE. Standard errors are clustered at the grant start year × field level.

Table A2: Over-requesting test

	Ratio
	(1)
$\%Over-request$	-0.001
	(0.001)
%NSF $Funding_i$	0.097***
	(0.017)
$\ln(Recommended_i)$	-0.028***
	(0.005)
Year FE	Yes
Field FE	Yes
Field × Year FE	Yes
Dep Var Mean	0.85
R-Sq	0.49
Observations	523

Notes: The unit of observation is a grant. For each grant, $\%Over - request$ is equal to $(Requested_i - Recommended_i)/Recommended_i$, which we use as a proxy for the share of over-requested resources. $\%NSF\,Funding$ is the share of funding coming from the NSF. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. Standard errors are clustered at the grant start year × field level.

Table A3: Correlation between papers retrieved only through the internal database and ratio

	(1) 1 {Paper only in XSEDE database}
$\%Allocated_i$	0.000
	(0.000)
$\ln(Recommended_i)$	0.027***
	(0.009)
%NSF $Funding_i$	-0.014
	(0.023)
Year FE	Yes
Field FE	Yes
Field × Year FE	Yes
Dep Var Mean	0.18
R-Sq	0.31
Observations	4824

Notes: The unit of observation is a paper. The outcome is an indicator equal to 1 if the paper is only reported in the internal XSEDE database. %*Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. $\%NSF\,Funding$ is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

Table A4: Matched vs Unmatched Grants

	$1\{Papers > 0\}$
	(1)
$\%Allocated_i$	0.060
	(0.126)
%NSF $Funding_i$	0.070*
	(0.036)
$\ln(Recommended_i)$	0.024*
	(0.014)
Field FE	Yes
Year FE	Yes
Field × Year FE	Yes
Dep Var Mean	0.68
R-Sq	0.38
Observations	844

Notes: The unit of observation is a grant. The dependent variable is an indicator equal to 1 if the grants is part of our sample of 711 grants and 0 otherwise. %*Allocated* is the share of resources allocated to a grant compared to the amount of resources recommended by the XRAC. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. $\%NSF\,Funding$ is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

B Tables: Main analyses using allocation levels

Table A5: Quantity of papers

	ln(papers)	asinh(papers)
	(1)	(2)
$\ln(Allocated_i)$	0.611***	0.564***
	(0.184)	(0.172)
$\ln(Recommended_i)$	-0.337*	-0.305*
	(0.176)	(0.164)
%NSF $Funding_i$	-0.071	-0.067
	(0.103)	(0.097)
Year FE	Yes	Yes
Field FE	Yes	Yes
Field × Year FE	Yes	Yes
Dep Var Mean	1.35	2.10
R-Sq	0.45	0.46
Observations	523	523

Notes: The unit of observation is a grant. In column (1), the outcome is the (log) number of papers associated with a grant. In column (2), the outcome is the inverse hyperbolic sine of the number of papers associated with a grant. The observation count shown is the count with remaining variation after fixed effects are included. ln(*Allocated*) is the log of the number of SUs allocated to the focal grant. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. $\%NSF\,Funding$ is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

Table A6: Use of concepts by others

	Log(Use)			Log(Use Prior)			Log(Use After)		
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(7) Mean	(8) Min	(9) Max
$\ln(Allocated_i)$	-0.433**	-1.225***	-0.245	-0.420**	-1.113***	-0.215	-0.502***	-1.539***	-0.230
	(0.186)	(0.311)	(0.210)	(0.194)	(0.318)	(0.216)	(0.189)	(0.323)	(0.199)
$\ln(Recommended_i)$	0.491**	1.101***	0.399*	0.457**	1.000***	0.345	0.559***	1.520***	0.401*
	(0.195)	(0.316)	(0.216)	(0.202)	(0.324)	(0.222)	(0.194)	(0.324)	(0.204)
%NSF $Funding_i$	-0.025	-0.083	-0.051	0.049	-0.017	0.025	-0.058	0.021	-0.101
	(0.120)	(0.145)	(0.137)	(0.103)	(0.144)	(0.122)	(0.133)	(0.145)	(0.148)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	-0.41	-1.27	0.07	-0.38	-1.20	0.10	-0.50	-0.99	0.05
R-Sq	0.67	0.55	0.63	0.68	0.57	0.64	0.69	0.63	0.65
Observations	465	397	465	462	383	462	448	309	448

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. In columns (1) to (3), we first calculate the average number of times concepts used in a paper are used by other PIs. Column (1) takes the mean of this value at the grant level, logged. Column (2) takes the min of this value at the grant level, logged. Column (3) takes the max of this value at the grant level, logged. Columns (4) to (6) repeat this exercise using the average number of times concepts used in a paper are used by other PIs prior to her entry into XSEDE. Columns (7) to (9) repeat this exercise using the average number of times concepts used in a paper are used by other PIs after her entry into XSEDE. ln(*Allocated*) is the log of the number of SUs allocated to the focal grant. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. %$NSF\,Funding$ is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

Table A7: Concept vintage

	Concept Year		
	(1) Mean	(2) Min	(3) Max
$\ln(Allocated_i)$	0.975	-0.294	2.808*
	(0.960)	(1.456)	(1.476)
$\ln(Recommended_i)$	-1.340	-0.947	-2.410
	(0.978)	(1.441)	(1.482)
%NSF $Funding_i$	-0.633	-0.664	-1.005
	(0.473)	(0.692)	(0.674)
Year FE	Yes	Yes	Yes
Field FE	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes
Dep Var Mean	2004	2000	2008
R-Sq	0.44	0.43	0.39
Observations	485	485	485

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. For each paper associated with a grant, we calculate the average year we observe the concepts it entails for the first time. Column (1) averages this measure at the grant level. Column (2) takes the min of this measure at the grant level. Column (3) takes the max of this measure at the grant level. ln(*Allocated*) is the log of the number of SUs allocated to the focal grant. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. %$NSF\ Funding$ is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

Table A8: Share of new concepts

	Vs prev. 2y			Vs prev. 3y			Vs prev. 5y		
	(1) Mean	(2) Min	(3) Max	(4) Mean	(5) Min	(6) Max	(7) Mean	(8) Min	(9) Max
$\ln(Allocated_i)$	0.040	-0.051	0.079**	0.056	-0.023	0.092***	0.052	-0.027	0.092***
	(0.037)	(0.061)	(0.034)	(0.038)	(0.070)	(0.032)	(0.040)	(0.069)	(0.034)
$\ln(Recommended_i)$	-0.043	0.007	-0.066*	-0.058	-0.019	-0.077**	-0.056	-0.019	-0.077**
	(0.038)	(0.062)	(0.034)	(0.039)	(0.070)	(0.033)	(0.041)	(0.069)	(0.034)
%NSF $Funding_i$	0.007	0.022	-0.006	0.000	0.012	-0.016	-0.001	0.014	-0.016
	(0.026)	(0.044)	(0.023)	(0.025)	(0.041)	(0.023)	(0.024)	(0.040)	(0.022)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Var Mean	0.78	0.63	0.88	0.76	0.61	0.87	0.74	0.59	0.86
R-Sq	0.45	0.37	0.42	0.45	0.37	0.43	0.45	0.37	0.44
Observations	430	430	430	438	438	438	446	446	446

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. For each paper, we calculate the share of 'new-to-the-researcher' concepts. Columns (1) to (3) compare XSEDE papers with papers written in the 2 years that precede XSEDE entry. Columns (4) to (6) compare XSEDE papers with papers written in the 3 years that precede XSEDE entry. Columns (7) to (9) compare XSEDE papers with papers written in the 5 years that precede XSEDE entry. Columns (1), (4) and (7) take the average of this measure at the grant level. Columns (2), (5) and (8) take the min of this measure at the grant level. Columns (3), (6) and (8) take the max of this measure at the grant level. ln(*Allocated*) is the log of the number of SUs allocated to the focal grant. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. %*NSF Funding* is the share of funding coming from the NSF. Standard errors are clustered at the grant start year × field level.

Table A9: Research Breadth

	Log(Unique Concepts)		
	(1) Mean	(2) Min	(3) Max
$\ln(Allocated_i)$	0.250**	-0.127	0.399***
	(0.110)	(0.224)	(0.122)
$\ln(Recommended_i)$	-0.263**	-0.128	-0.359***
	(0.109)	(0.223)	(0.122)
%NSF $Funding_i$	0.042	0.155	0.021
	(0.046)	(0.124)	(0.054)
Year FE	Yes	Yes	Yes
Field FE	Yes	Yes	Yes
Field × Year FE	Yes	Yes	Yes
Dep Var Mean	3.09	2.32	3.37
R-Sq	0.35	0.42	0.35
Observations	475	475	475

Notes: The unit of observation is a grant. The observation count shown is the count with remaining variation after fixed effects are included. We calculate for each paper the number of concepts it entails. Column (1) averages this measure at the grant level, logged. Column (2) takes the min of this measure at the grant level, logged. Column (3) takes the max of this measure at the grant level, logged. ln(*Allocated*) is the log of the number of SUs allocated to the focal grant. ln(*Recommended*) is the log of the number of SUs recommended by XRAC. %*NSF Funding* is the share of funding coming from the NSF. All columns control for the amount recommended. Standard errors are clustered at the grant start year × field level.

C XSEDE Review & Reconciliation Process

This section provides additional details about the XSEDE Review and Reconciliation process, through which recommendations are made and then reduced to final allocations. In addition to conversations with individuals at XSEDE, much of the following (including excerpts) comes from the 2021 v1.1 versions of *XSEDE Allocations Practices and Procedures* and *XRAC Reviewer Manual*.[26]

C.1 Review Criteria:

The excerpt below from the *XRAC Reviewer Manual* summarizes the criteria reviewers should consider when making recommendations from proposals:

1. *Appropriateness of Methodology: Does the request describe appropriate tools, methods, and approaches for addressing the research objectives? These methodologies may be community codes or models, data analysis methods, or algorithmic formulations expressed in user-developed scripts or tools.*
2. *Appropriateness of Research Plan: Does the request describe necessary and sufficient computational experiments to answer the research questions posed? In some cases, the research plan may be more reasonably expressed as estimates of resource use, supported by past or early experience. Serious concerns about the research plan will be documented in reviews and may lead to reduced allocation awards.*
3. *Efficient Use of Resources: Has the request identified appropriate resources for undertaking the research plan with the proposed methodology? And will those resources be used as efficiently as is reasonably possible and in accordance with the recommended use guidelines of those resources? Is the proposed work better suited or equally well suited to other resources available for allocation?*

XRAC reviewers may also consider these "additional considerations:"

> *Intellectual Merit: Reviewers also consider whether the work has any supporting grants for which the science has been merit-reviewed. If so, is the allocation request consistent with the objectives of the supporting grants as described in the submission and is the scale of resource use commensurate with the level of support? If no such grants are identified, the reviewers must assess the intellectual merit of the work. This assessment of the intellectual merit will factor into their overall recommendation.*

In our context, the intellectual merit consideration serves to put all requests on a "level playing field." The vast majority of grants are supported by government agencies. Those that are validated to be sure the science proposed is legitimate.

[26]The reviewer manual can be found at https://www.ideals.illinois.edu/items/111788/bitstreams/366207/data.pdf. The practices & procedures document has been removed from the internet but can be shared by the authors upon request.

Prior results: Research requests typically require some form of prior work or progress to be considered for an allocation. For New requests, demonstrated benchmarking or test runs are typically required on the requested resource in almost all cases; in some instances (such as a request for a new resource for which Startups are not yet available), demonstrated work on an architecturally similar resource may suffice. For Renewal requests, reviewers consider the provided Progress Report as well as utilization levels of prior allocations

The prior review criteria ensures researchers are using their existing allocations (i.e., are not wasteful). In conversations with XSEDE directors, this criteria serves primarily to check that prior allocations are indeed "doing something." In particular, reviewers may look at if the project has produced *any* publications. Hence, if this leads some values of *recommended* to overestimate the *true need* for the more productive projects, one might worry that this may introduce bias to our quantity (number of papers) analysis. Our data does not allow us to empirically address this threat because we cannot confirm if unmatched grants truly produce zero papers or do not cite XSEDE (see Section 5), this would in fact bias our estimates in Table 4 downward causing us to underestimate the effect. This can be seen in a simple numerical example. Suppose we observe a project who had already published a paper using a previous XSEDE allocation and the PI requests a renewal. This project is recommended 100 SU and is awarded 50 SU ($\%Allocated = 50\%$). If, in reality, this project had not already published a paper, the project may have been recommended 90 SU. Because larger recommendations are reduced more (see reduction formula below), this project will be awarded an allocation greater than 45 SU (such that $\%Allocated > 50\%$). Hence, the project with prior publications (and higher expected productivity) has a lower ratio than it would and a higher output – biasing the results downward.

C.1.1 XRAC Recommendations and Resource Availability

The excerpt below illustrates that reviewers are instructed to make recommendations independent of available XSEDE resources.

In all parts of the process, reviewers are asked to consider requests only on their merits. The reviewers are not informed about the total requests on the different systems until the caucus session and only after their individual reviews have been completed. This approach is designed to ensure that reviews are based purely on the merit of the proposal. Providing the allocation request total and available units for each resource at the caucus allows them time to prepare mobility information for requests that need to be or could possibly be moved to alternate systems in an oversubscription situation. The total of the Recommended Allocations is reported to NSF, so that NSF will be aware of the panel's statement of the actual need for resources.

C.2 Reconciliation formula

After recommendations have been made by the XRAC for each proposal, the Allocations team meets to "reconcile" recommendations with the resources available. The first phase is called a "balancing" phase, where allocations from over-recommended resources are moved to architecturally equivalent under-subscribed resources until the excess demand is balanced equally across computing platforms. If there is still over-subscription after this phase, the balancing phase is followed by a "reconciliation" phase where recommendations are reduced through a formulaic solver in order to determine allocations. The formula gives priority to NSF-funded projects.

The reconciliation formula is the following:

$$\text{Award} = [(1 - G) \times R_n + (1 - G) \times F \times R_o] \times Size(R, S)$$

with G the global scaling factor that is solved for by non-linear optimization, R the Recommended amount, R_n the Recommended amount supported by NSF-funding, R_o the Recommended amount non-supported by NSF-funding, F the funding priority factor and S a size-scaling factor (smaller awards are reduced less).

The process is described in the practices and procedures document as:

> *Following the discussion of all requests, if the aggregate Recommended Allocation is less than the total Available Amount for a system, the Recommended Allocations generally become the Awarded Allocations. When the total Recommended Allocation exceeds the total Available Allocation for a given resource, XSEDE and SP staff use a procedure approved by the NSF to adjust the Recommended Allocations to arrive at the Awarded Allocations. This Reconciliation adjustment is performed by the XSEDE Allocations staff and SP representatives after the XRAC meeting ends and formulaically adjusts all Recommended Allocations to remove oversubscription. Within two weeks of the meeting, the XSEDE allocations staff complete the Reconciliation process, summarize the panel discussions, craft administrative comments or notes to PIs as needed, and send notifications of final awards to all PIs, co-PIs and Allocation Managers of requesting projects, as well as to all affected SPs.*

D Concepts

Definition

Dimensions AI categorizes and organizes vast amounts of academic research data using machine learning techniques. One of its key features that we utilize in this paper is its ability to extract and identify "concepts" from scientific texts.

Concept Identification: Dimensions AI scans through publications' abstracts and identifies "concepts" associated with each paper. These concepts are meant to represent key themes or topics within a paper. They are derived from the text using machine learning models that recognize important keywords, phrases, and themes commonly associated with specific scientific domains.

Relevance Scoring: Each concept identified by Dimensions is assigned a relevance score, which indicates how central the concept is. While the exact scoring scheme is proprietary, it is a function of the field of study of the paper.

Differences with other methods: The algorithm of Dimensions AI differs from other algorithms such as LDA where the number of topics has to be imputed. This allows us to have variation in breadth. In that sense, Dimensions AI concepts are closer in idea to Hierarchical Dirichlet Process (HDP) that have been used previously in the literature, where the number of topics assigned to a document is not pre-determined.[27]

We present two examples to illustrate the use of concepts. The first paper is relatively narrow in scope, with only four concepts. The second paper is broader, with 22 concepts.

Paper 1: Defect-engineered graphene chemical sensors with ultrahigh sensitivity

- Abstract: *We report defect-engineered graphene chemical sensors with ultrahigh sensitivity (e.g., 33% improvement in NO2 sensing and 614% improvement in NH3 sensing). A conventional reactive ion etching system was used to introduce the defects in a controlled manner. The sensitivity of graphene-based chemical sensors increased with increasing defect density until the vacancy-dominant region was reached. In addition, the mechanism of gas sensing was systematically investigated via experiments and density functional theory calculations, which indicated that the vacancy defect is a major contributing factor to the enhanced sensitivity. This study revealed that defect engineering in graphene has significant potential for fabricating ultra-sensitive graphene chemical sensors.*
- https://doi.org/10.1039/C5CP04422G
- Concepts: chemical sensors, conventional reactive ion, density functional theory calculations, ultrahigh sensitivity
- Breadth (as calculated with our measure): 4

Paper 2: Hybrid massively parallel fast sweeping method for static Hamilton–Jacobi equations

[27] See for instance Furman & Teodoridis (2020).

- Abstract: *The fast sweeping method is a popular algorithm for solving a variety of static Hamilton–Jacobi equations. Fast sweeping algorithms for parallel computing have been developed, but are severely limited. In this work, we present a multilevel, hybrid parallel algorithm that combines the desirable traits of two distinct parallel methods. The fine and coarse grained components of the algorithm take advantage of heterogeneous computer architecture common in high performance computing facilities. We present the algorithm and demonstrate its effectiveness on a set of example problems including optimal control, dynamic games, and seismic wave propagation. We give results for convergence, parallel scaling, and show state-of-the-art speedup values for the fast sweeping method.*
- https://doi.org/10.1016/j.jcp.2016.06.023
- Concepts: Hamilton-Jacobi equation, algorithm, computer architecture, dynamic game, equations, example problems, fast sweeping method, heterogeneous computer architectures, high performance, hybrid parallel algorithm, optimal control, parallel algorithm, parallel computing, parallel method, parallel scaling, popular algorithms, seismic wave propagation, speedup values, static Hamilton–Jacobi equations, sweeping algorithm, sweeping method, wave propagation
- Breadth (as calculated with our measure): 22