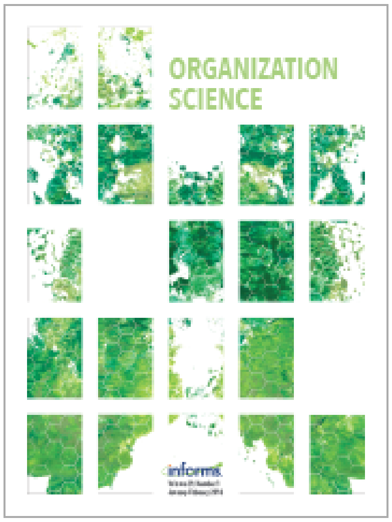


Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering

Jeffrey L. Furman, Florenta Teodoridis

informs
http://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Vol. 31, No. 2, March–April 2020, pp. 330–354
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering

Jeffrey L. Furman,[a,b] **Florenta Teodoridis**[c]

[a] Strategy & Innovation Department, Boston University, Boston, Massachusetts 02215; [b] National Bureau of Economic Research (NBER), Cambridge, Massachusetts 02138; [c] Department of Management and Organization, Marshall School of Business, University of Southern California, Los Angeles, Los Angeles, California 90089

Contact: furman@bu.edu, http://orcid.org/0000-0002-6273-8380 (JLF); florenta.teodoridis@marshall.usc.edu, http://orcid.org/0000-0002-4917-5552 (FT)

Received: August 2, 2018
Revised: February 7, 2019
Accepted: April 20, 2019
Published Online in Articles in Advance: January 31, 2020

https://doi.org/10.1287/orsc.2019.1308

Abstract. We examine how the introduction of a technology that automates research tasks influences the rate and type of researchers' knowledge production. To do this, we leverage the unanticipated arrival of an automating motion-sensing research technology that occurred as a consequence of the introduction and subsequent hacking of the Microsoft Kinect system. To estimate whether this technology induces changes in the type of knowledge produced, we employ novel measures based on machine learning (topic modeling) techniques and traditional measures based on bibliometric indicators. Our analysis demonstrates that the shock associated with the introduction of Kinect increased the production of ideas and induced researchers to pursue ideas more diverse than and distant from their original trajectories. We find that this holds for both researchers who had published in motion-sensing research prior to the Kinect shock (within-area researchers) and those who did not (outside-area researchers), with the effects being stronger among outside-area researchers.

Funding: This work was supported by the National Science Foundation's Program on the Science of Science and Innovation Policy (SciSIP) [Grants SES-1564368 and SES-1564381].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2019.1308.

Keywords: automation • knowledge production • innovation • research technology • rate and direction of innovation • technological change • topic modeling • machine learning • idea space • research trajectories • knowledge trajectories • diversification • breadth and depth of knowledge

1. Introduction

The automation of physical and mental tasks in the production of goods and services is among the most profound factors affecting the modern economy and the role of human capital in economic output (Mokyr 2002). Recent work on this topic documents that the extent to which automating technologies substitute for or complement human labor depends on the extent to which such technologies substitute for the specific tasks associated with production (Deming 2017, Acemoglu and Restrepo 2018, Brynjolfsson et al. 2018, Felten et al. 2018). However, to date, most of the work on this topic has focused on manufacturing or service industries, whereas relatively less research examines the impact of such automating technologies on the production of new-to-the-world knowledge (Ding et al. 2010, Cockburn et al. 2017).

In this paper, we contribute to investigating the impact of information technology (IT)–based automating technology[1] on the production of knowledge. This is important because of the central role knowledge production plays in economic growth and the unique features of the knowledge production environment that may provide insights into the role of automation in other sectors. The impact of automating technologies on knowledge production may differ in important ways from the impact of such technologies on goods and services production because of the variety of tasks associated with knowledge work and the extent of autonomy that knowledge workers possess in allocating time across such tasks. These features may enable knowledge workers to respond to automating technologies in ways that differ from those of sectors that face greater rigidity in task reallocation and, potentially, a greater chance for displacement of tasks previously performed by humans (e.g., Acemoglu and Autor 2011). Furthermore, in knowledge production, tasks are knowledge based in the sense that the completion of a research task requires a certain type of research expertise. Hence, we anticipate that researchers' response to automating technologies will be influenced by the overlap in knowledge that is being automated through the research technology (Murray et al. 2016, Thomson and Zyontz 2017, Teodoridis 2018). As a result, we distinguish between two types of researchers: those who had previously worked in the research area where the tasks affected by the automating research technology reside, whom we call *within-area researchers*, and those who had not previously worked in the area, whom we describe as *outside-area researchers*.

Estimating a relationship between the availability of automating research technology and the rate and type of knowledge output is difficult because numerous unobservable factors are likely associated with both research technology availability and researchers' project choices. To generate causal insights, we exploit an unexpected arrival of research technology as an instrument that is correlated with research task automation but not with factors affecting researchers' efforts, other than through its effect on automation. Building on Teodoridis (2018), we leverage the introduction of the Microsoft Kinect gaming system as an unexpected shock that suddenly and dramatically made available a technology that automates core motion-sensing research tasks. Kinect was launched in November 2010 as an add-on for Microsoft's Xbox 360 gaming console. Soon after its launch, hackers developed and released through the open-source community a driver that enabled devices other than the Xbox to interact with Kinect, thus making it possible for scholars in electrical engineering, computer science, and electronics to harness Kinect's motion-sensing output for use in research applications.

Kinect automates several key tasks required to track, collect, and analyze real-time complex three-dimensional (3D) motion data. Prior to Kinect, in order to complete these tasks, researchers had to rely on developing complex custom algorithms for collecting, analyzing, and attempting to augment the output of two-dimensional (2D and 3D cameras. It is important to note that full automation of motion-sensing research tasks implies a system that can recognize and understand any 3D motion scene. Whereas such a system is still to be developed, Kinect helped move a large step closer toward it (Han et al. 2013). In our email exchanges with Kinect researchers, one individual noted that

> [with Kinect], researchers can get [motion] data without worrying about how it is obtained in the most of indoor scenes. Now, the research topics have been moving from how to get to shape and motion to how to use that information for applications such as robots and artificial intelligence.

We use the Kinect shock in difference-in-differences estimations to ascertain the impact of its introduction on measures of researcher-level project choice in the domains of computer science and electrical and electronics engineering research. To measure researchers' rate and type of knowledge output, we rely on bibliometric indicators. Counts of academic publications and citations are established (though imperfect) measures of research output. We use these to evaluate the impact of Kinect on researchers' rate of knowledge output (i.e., productivity). Estimating changes in the type of knowledge output presents a greater challenge because it requires tracking research trajectories in an ever-evolving multidimensional space of ideas. We focus on two characteristics of the type of knowledge output: changes in the diversity of research projects and changes in the trajectory of research.[2] We propose novel measures of units of knowledge in idea space based on topic modeling analysis using unassisted machine learning techniques that capture the latent categorization of academic publications without relying on an ex ante preimposed structure (such as author-defined keywords or institutionally defined research fields). Our analysis also considers more traditional measures of research diversity and trajectories based on observable characteristics of academic publications. Each measure involves advantages and disadvantages that we discuss in our analysis. Although each measure captures somewhat different attributes of research behavior, we believe that taken together they provide an informative window into the effects of automating research technology on researchers' type of knowledge output.

Contrary to concerns that automation displaces human labor and consistent with the idea that knowledge workers face substantial latitude for shifting across tasks, we find that the automation of key research tasks leads both within-area researchers and outside-area researchers to experience (1) an increase in research output, (2) an increase in the diversity of their research, and (3) a shift in their research trajectories. The outside-area researchers experience the largest impact, whereas within-area researchers experience a more modest effect. Interestingly, outside-area researchers achieve a boost both in their work on motion-sensing research and in the set of their projects that do not directly engage with motion-sensing. For example, an outside-area researcher in our sample who, prior to Kinect, focused on research involving sound waves extended his research to the study of infant seizures by developing detection techniques combining audio and video inputs. Among the group of within-area researchers, those whose research was most focused on motion-sensing before Kinect experience the highest benefits. This suggests that the automating research technology allows within-area researchers to work more efficiently and to engage in new types of projects. For example, after Kinect, a within-area researcher in our sample engaged in adapting computer vision detection and visualization algorithms to developing malaria diagnostic and tumor identification techniques, whereas another researcher pushed forward her motion-sensing agenda to include studies of virtual reality. Although these results are consistent with our hypotheses, they are not obvious ex ante.

Automating technologies often increase returns for incumbents relative to entrants and do not necessarily enable organizational diversification.

Our analysis underscores the significant role played by research technology in knowledge production and sheds light on the processes by which the presence of technologies that automate certain research tasks influences knowledge workers' behavior. Understanding these processes is thus important for the study of the strategic management of science-based organizations (Murray and O'Mahony 2007, Nelson 2016), including research-oriented private firms competing for profits and research institutions competing for reputation and opportunities for knowledge creation. Furthermore, our study suggests that firms' decisions to engage in technology development and their technology commercialization strategy might influence knowledge flows and hence subsequent innovation efforts.

2. Automating Technology and Knowledge Production

2.1 Automating Technology in Knowledge Production vs. Other Economic Sectors

While relatively little scholarly work has examined the impact of automating technology on knowledge production, a substantial amount of work has been devoted to understanding the impact of such technological progress on the production of physical goods and, more recently, services. Since the earliest days of the industrial revolution, concerns about the impact of automating technology on labor and society have been a recurrent feature of public and academic debate (Mokyr et al. 2015). Over the past few decades, research in management and economics has examined the impact of technology on multiple outcomes, including firm-level outcomes such as productivity (Pinsonneault and Kraemer 2002, Brynjolfsson and Hitt 2003), competitive performance (Dos Santos and Peffers 1995, Bloom et al. 2012), organizational change (Dean et al. 1992, Argyres 1999, Volkoff et al. 2007), and firm boundaries (Forman and McElheran 2012). More recently, scholars have begun to study the economic impact of machine learning on people, firms, and society (Agrawal et al. 2018, Choudhury et al. 2018).

A persistent question in this literature regards whether automating technologies constitute substitutes or complements for human labor. Most recently, the focus of the literature has shifted from approaching this question with a view of whether technology and labor are substitutes overall and toward the recognition that the impact of automation on labor depends on the extent to which technology substitutes for specific tasks and affects the allocation of human effort across tasks within a personnel function (Deming 2017, Acemoglu and Restrepo 2018, Brynjolfsson et al. 2018, Felten et al. 2018). This question is important for both macro-level issues, such as the share of labor in the economy (Karabarbounis and Neiman 2014), and micro-level issues, such as the organization of work and the level of firm productivity (Bresnahan et al. 2002, Bloom et al. 2012).

An important underlying assumption in tackling this question is that automation is a process that reduces the cost of performing certain tasks. It is this cost reduction that creates incentives for economic actors to substitute automating technology for human labor (Acemoglu and Autor 2011, Nordhaus 2007). Thus, when automating technology improves the productivity of a task, people and organizations shift human labor away from that task and, under certain conditions, onto other tasks.[3]

In knowledge production, we expect similar forces at play but in a context where workers have substantial autonomy in choosing their tasks and a broad array of tasks to choose from. As Smith (1776) observed in his pin factory visits, manufacturing jobs can often involve substantial task specialization. In addition, manufacturing jobs characterized by task specialization are often embedded in relatively rigid hierarchies that provide relatively little worker autonomy. By contrast, the job of knowledge production, particularly for academic researchers, involves both a broad array of potential tasks and substantial autonomy (Aghion et al. 2008, Cohen et al. 2018). The tasks of a researcher are often extremely varied and can include project conception, choice of project partners and coauthors, grant writing, choice of technical and support staff, and project execution across a variety of dimensions (e.g., preparation of research materials, conduct of specific analyses, project presentations, conference attendance, collaboration with industry). Further, research autonomy, including the freedom to choose research topics and to choose how to allocate time across a variety of research tasks, is the hallmark of scientific scholarship (Merton 1938). Although many service professions also involve a wide breadth of tasks and enable some autonomy (e.g., graphic design, civil engineering, legal practice, investment banking, financial consulting), others have a lesser degree of both (e.g., bank tellers, hair stylists, fast-food workers, financial accountants). In other words, we consider knowledge production to be an area in which task variety and autonomy are particularly great and thus a sector worthy of attention, especially considering the impactful role of knowledge production in the economy.

2.2. Automating Research Technology and Research Expertise

We hypothesize that technology automation will differently affect researchers with substantial expertise in the research area into which the automating technology becomes available, that is, within-area researchers, than those who do not have specialized human capital in that area, that is, outside-area researchers. The case of statistical analysis packages provides an illustrative example. Statistical software such as SAS, Stata, and R automates tasks that would otherwise require substantial human capital investments, including merging data, computing test statistics, inverting matrices, and composing graphics. Researchers could complete these tasks if they made sufficient investments in their human capital. Alternatively, researchers could collaborate with people who have made such human capital investments. The ready availability of off-the-shelf software, however, automates each of these tasks and, via revealed preference, appears to enable researchers to complete their work more efficiently than would be possible in the absence of such software. In addition, the software packages may enable researchers to shift the time made available by the automation of statistical analysis to other research activities, including undertaking new and different projects. Although these effects likely hold for all researchers who conduct quantitative analysis, they likely differ in their impact and nature for experts in statistical analysis, for example, econometricians, relative to other researchers. Econometricians might be displaced from their core research tasks and that could lead to exit or to additional time to be spent on other projects such as advancing the field of statistical analysis. For other researchers, the availability of the technology could enable them to employ the technology as a substitute for performing the task using specialized human capital.

As this example suggests, the automation of some core research tasks could affect the knowledge production function in several ways and may do so differently for researchers who have developed specialized human capital related to the now-automated tasks and for those who have not. We formalize these thoughts in a set of hypotheses below.

2.2.1. Automating Research Technology and Within-Area Researchers. We begin by focusing on researchers with historical expertise in the domain affected by the automation. The automation of core research tasks can have multiple effects for these people. If a researcher's entire efforts are focused on the area in which the automation occurs, it is possible that the automation will substitute for his or her efforts, and his or her productivity will decline or he or she will exit the domain of research or the conduct of research entirely. This would be consistent with Autor et al. (1998), who highlight the prospect that technology substitutes for related human capital. If, however, the automated tasks constitute a consequential fraction of the researcher's tasks but not the full set of those tasks, he or she may reallocate his or her time to these other tasks. Thus, because the productivity of the automated tasks rises while the cost of the other tasks remains constant, we expect the productivity of such researchers to improve.[4] Considering the variety of tasks associated with research and the extensive autonomy of most researchers, we anticipate a positive net effect of automating technology for within-area researchers.

Hypothesis 1a. *The availability of research technology that automates core research tasks boosts the rate of research output of within-area researchers.*

An additional first-order effect of research task automation regards its influence on researcher's type of knowledge output. Aghion et al. (2008) suggest that research technology that affects the productivity of domain-specific knowledge may induce greater mobility across research trajectories. The idea is that the increase in productivity of the automated task allows these within-area researchers to spend the additional time on either experimenting with new topics or applying their expertise to projects in other domains. Thus, the availability of automating research technology may boost the opportunities set among within-area researchers, facilitating tackling new research questions within their domain of expertise or branching out into other research trajectories.

Hypothesis 1b. *The availability of research technology that automates core research tasks changes the composition of research project types for within-area researchers.*

2.2.2. Automating Research Technology and Outside-Area Researchers. Researchers who have not invested in the specialized human capital associated with the tasks automated by the research technology can be affected by the automation in ways that differ from those of within-area researchers. A core insight here is that an automating research technology that reduces the cost of performing certain research tasks compensates for the fixed costs of research expertise within the knowledge area of the technology (Cohen and Klepper 1992, 1996). This, in turn, reduces returns to that within-area knowledge and, hence, opens opportunities for outside-area researchers. For example, Furman and Stern (2011) and Murray et al. (2016) suggest that open access to research tools can exert a democratizing influence on other fields and may compensate for the fixed costs of within-area expertise. In addition, outside-area researchers' lack of specialized skills in the affected domain means that they are less likely to experience substitution in any of their research tasks. Thus, adopting the automating technology will have a positive impact on researchers' research productivity. Furthermore, the impact could be higher than that for within-area researchers who experience a boost in productivity through some degree of task substitution.

Hypothesis 2a. *The availability of research technology that automates core research tasks boosts the rate of research output of outside-area researchers on all topics on which they work. The effect is greater than for within-area researchers.*

In addition to affecting productivity, we anticipate that the automating technology will influence outside-area researchers' type of knowledge output. Teodoridis (2018) shows that a democratizing change in the availability of research technology alters team composition to include specialists from knowledge areas outside the domain of the research technology. To the extent that changes in collaboration reflect changes in researchers' project choices, this suggests the possibility of a change in the project portfolio composition of outside-area researchers. Moreover, we expect that the automating research technology will influence the composition of research project types of outside-area researchers by creating opportunities for novel idea recombinations (Uzzi et al. 2013) and, thereby, new lines of inquiry. Thus, we expect that outside-area researchers can leverage the availability of automating research technology to explore ideas that incorporate the automated tasks either directly in the research domain of the technology or by broadening lines of inquiry in their areas of historical focus. As before, we expect the impact to be higher than that for within-area researchers, who also experience a change in the opportunities they face but through some degree of task substitution.

Hypothesis 2b. *The availability of technology that automates core research tasks changes the composition of research project types of outside-area researchers. The effect will be greater than for within-area researchers.*

3. The Kinect Shock and the Automation of Motion-Sensing Research

The shock we examine in this paper is the launch of the Microsoft Kinect gaming system that suddenly and dramatically made available a research technology that automates certain motion-sensing research tasks. We exploit the unexpected availability of this technology as an instrument that is correlated with task automation in knowledge production but not with factors affecting researchers' efforts, other than through its effect on automation.

The setting appears unusual at first glance: it is the result of the unexpected impact of Microsoft's successful launch of a controller-free video game system designed to compete with rival products launched by Nintendo and Sony. In the two months following Kinect's launch on November 4, 2010, Microsoft sold more than 8 million units (>130,000 Kinect units per day), outpacing the iPhone and the iPad to become the Guinness World Records' all-time fastest-selling consumer electronic device (Bilton 2011). The surprise that makes Kinect valuable for our research context is not, however, its commercial success but its wide-ranging and near-immediate impact on scholarship in some areas of computer science and engineering.

3.1. Microsoft's Introduction of the Kinect System

On November 4, 2010, Microsoft introduced the Kinect system for its Xbox video game console with the aim of competing with handheld gesture-recognition remotes introduced previously by Nintendo (Wii) and Sony (PlayStation). With the Kinect, Microsoft attempted to leapfrog its video console rivals by creating the first hands-free controller for electronic devices, a game controller system that responded to the natural movements of the player.[5]

In addition to being a treat for video gamers, Kinect also constituted a feast for hackers, who descended on the system with the aim of accessing the vast data obtained by Kinect's sensors and linking them to other devices, such as computers and robots. These efforts received a twofold infusion of interest on Kinect's launch day. The first came from Adafruit Industries, a manufacturer of do-it-yourself electronics kits operated by alumni of MIT's Media Laboratory, Limor Fried and Phillip Torrone, that offered a $1,000 prize for the first individual or organization to post an open-source Kinect driver to GitHub (Carmody 2010a). The second spur of interest arose as a result of Microsoft's active (and quixotic) effort to thwart the hackers. In a same-day response to Adafruit's prize offer, Microsoft released a statement to CNet: "Microsoft does not condone the modification of its products . . . [and will] . . . work closely with law enforcement and product-safety groups to keep Kinect tamper-resistant" (Terdiman 2010, p. 1). Adafruit responded immediately by doubling its Kinect driver bounty to $2,000, further intensifying the race for the driver.

The race was won on November 11 by a Spanish computer science undergraduate student, Héctor Martín, who did not own an Xbox but who had purchased a Kinect that morning when it went on sale in Europe (Giles 2010). Within days of the driver's release, researchers and hobbyists had adapted Kinect for numerous uses, including the creation of 3D computer holograms and a modified iRobot Roomba that could respond to human hand and voice commands and could create visual maps of the rooms it had visited (Wortham 2010).

During the week that hackers had raced to create an open-source driver to harvest Kinect's data, consumers purchased nearly a million Kinect units. In the wake of its success, Microsoft initially continued to resist working with the hacker community and even refused to acknowledge that its system had been hacked. Within 10 days of the release of Martín's open-source driver, however, Microsoft, convinced of the value of embracing the experimentation of the hobbyist and scientific communities, pivoted entirely and announced that it would not pursue legal remedies against those who adapted the Kinect system for other purposes (Carmody 2010b).

3.2. Kinect as an Automating Research Technology

Motion-sensing research involves a series of topics in the broader research area of computer vision, an interdisciplinary field in computer science and electrical engineering with the principal goal of enabling machines to "see" as well as (or better than) humans. Achieving this goal requires the ability to scan 3D environments and recognize where objects appear in both static and dynamic environments and under various lighting conditions. While often taken for granted by humans, vision is an exceptionally complex task for machines. Humans use their eyes to observe 3D scenes and their brains to process the types of objects and movement in the scenes. Machines rely on cameras as their "eyes" and on human developed software and hardware as their "brain."

Prior to the introduction of Kinect, there were two main approaches to motion-sensing in computer vision. In a first generation, researchers employed 2D cameras and developed software as the sole method of inferring depth and movement based on the images captured by those cameras. This process was complex and time consuming because it required sophisticated mathematical calculations, such as those required to infer depth based on known object sizes and movement from serial images. Such custom-developed software was not particularly fast to run, and it was prone to errors because it relied heavily on factors such as the resolution of the images, the frequency of frames, and the ingenuity of the software developer in coding efficient calculations. The second generation of this work involved advances in images captured from 2D cameras to 3D cameras, such as time-of-flight cameras. These early 3D cameras operated with low resolution, were subject to especially high sensitivity to lighting conditions, and were not particularly accurate in tracking movements across all three dimensions. Although the availability of such cameras lessened the need for some of the coding development tasks required to process the image output, a broad set of such tasks remained even after these cameras were introduced, including the needs to clean image data, accurately trace movement, speed up the processing of the image output, infer missing pieces in various unfavorable lighting conditions, and impute missing information resulting from low-resolution images. Whereas some algorithms for working with these cameras and their output could be shared, the vast majority needed to be customized for the environment in which individual researchers were working.

By offering higher-resolution 3D images and an embedded processing capability that more accurately traced movement under a variety of lighting conditions, Kinect represented a significant technical advance that eliminated a substantial set of coding tasks that had previously fallen on motion-sensing researchers. The advance enabled by Microsoft's new tool was sufficiently great that it allowed a novice to extract and use Kinect's output without the need for specialist coding to process the data to render them useful. As a result, Kinect automated many tasks that previously required specialized humans involved in developing software algorithms to process images captured by cameras with a goal of uncovering the same insights that a human would when observing a 3D scene in movement. Kinect did not fully resolve all challenges associated with computer vision; for example, Kinect performed poorly in bright sunlight but automated a substantial set of tasks that had previously required specialized human coding.

The impact of Kinect was perceived by the computer vision research community to be wide ranging. For example, a computer vision researcher told us that "Kinect had a game-changing effect on the research possibilities. We work in robotics perception, that is, how can robots perceive and act in the environments. Because our world is 3D and Kinect gives 3D information, the data become extremely powerful. This has enabled significant advances in applications such as object detection, human activity recognition and anticipation for robots, as well as robotic grasping and path planning." In addition to its impact in computer vision research, Kinect raised the interest of researchers from other domains. Richards-Riessetto et al. (2012) describe the value of Kinect for work in archaeology, and Rafibakhsh et al. (2012) describe its value for construction engineering and management. In our own discussions with motion-sensing researchers, a researcher noted that his group provided a Kinect-based algorithm developed in his laboratory to

> ICT's Medical VR group which applied [it] to various motor rehabilitation applications. We've also researched the use of Kinect for human activity analysis (examining body language, fidgeting, etc.) to help therapists understand behavior of their patients. . . . Our lab's director and some of his students used the Kinect to create really interesting education game experiences. We have put together a number of other experiments, prototypes,

> and demonstrations involving the Kinect. It runs the gamut from wide area tracking to virtual human puppeteering to 3D scanning, and more. I can't list them all.

More broadly, after the launch of Kinect, motion-sensing appears to have found its way into an increasing variety of research projects with applications in a wide set of domains, from artificial intelligence and virtual reality to education, healthcare, music, cinematography, market research, and advertising. For example, faculty and graduate students at MIT's CSAIL laboratory have designed a motion-sensing system, called Emerald, that tracks peoples' movements within their homes, can alert medical personnel in the event of a medical catastrophe or fall, and can even be used to predict fall events.

In addition to reducing the need for specialized human capital investments, a secondary impact of Kinect is that it also reduced the monetary cost of capturing and leveraging data from 3D images. As we noted earlier, an important underlying aspect of automation that explains the incentives it generates for economic actors to substitute the technology for human labor is that automation is a process that reduces the cost of performing certain tasks with human labor (Acemoglu and Autor 2011, Nordhaus 2007). While Kinect indeed reduced the cost of executing certain research tasks with human labor, for example, software coding, the monetary cost of Kinect was lower than that of the previous generation of 3D cameras: whereas the cost of time-of-flight cameras ranged from $10,000 to $20,000, Kinect cost $150 at launch. To sharply isolate the reduction in cost via automation, we would have needed for Kinect to be priced in the same range as the previous generation of motion-sensing technology (e.g., time-of-flight cameras). While this remains a limitation of our study, we believe that the dominant effect is that of the reduction in cost through automation. The reason is that for researchers committed to computer vision research (within-area researchers), this change in monetary cost was likely not especially great relative to the overall cost of operating their laboratories. In other words, we argue that the Kinect impact would have been roughly the same even if Kinect was to be priced in the same range as the previous generation of motion-sensing technology. For outside-area researchers, however, the change in cost may have enabled experimentation that these researchers would not have considered if Kinect were priced in the same range as time-of-flight cameras. It may be, therefore, that changes in the monetary cost of motion-sensing research technology drive some of the effects we observe. We believe, however, that the more substantial change enabled by Kinect is not related to the price of the research technology but is more related to the fact that the advent of Kinect obviated the need for specialized human capital and enabled researchers in other areas to appreciate the potential value of motion-sensing techniques to sets of problems related to their areas of interest.

Overall, the launch of Microsoft Kinect appears to have changed the opportunity set for innovation in computer science and engineering. This development appears to be exogenous and to have been a surprise to the incumbent research community, the community of potential users that had been working outside traditional motion-sensing topics, and even to Microsoft itself.

4. Data and Empirical Strategy

To examine the impact of Kinect on researchers' productivity and portfolio of project types, we draw on the population of publications, early-access publications, and conference proceeding papers included in the Institute of Electrical and Electronics Engineers' (IEEE) Xplore database, which covers nearly 200 computer science and electrical engineering journals and more than 1,800 conference proceedings, between 2001 and 2014.[6] We conduct our estimation on the subset of papers published in the four years before and four years after the launch of Kinect (2007–2014), and we use the remainder of the data to obtain better estimates of researchers' pre-Kinect research behavior and trends.

4.1. Empirical Strategy

We employ a difference-in-differences analysis to compare research productivity and type of knowledge output before and after the launch of Kinect. Formally, we estimate for researcher i and year t

$$DV_{it} = \beta(TreatedResearcher_i \times AfterKinect_t) + Age_i + Age_i^2 + \delta_i + \gamma_t + \epsilon_{it}, \tag{1}$$

where $TreatedResearcher_i$ is a dummy variable equal to 1 if research i is a treated unit and 0 otherwise. We define treated researchers as individuals who were publishing in motion-sensing before the launch of Kinect or academics who started to publish in motion-sensing only after the launch of Kinect. To identify motion-sensing publications, we search the full text and metadata of publications in the IEEE Xplore database using carefully identified keywords and through interviews with subject-matter experts and cross-referenced with IEEE's taxonomy.[7] The term $AfterKinect_t$ is a dummy variable equal to 1 if the observation year is between 2011 and 2014, namely after Kinect's launch, and 0 otherwise. We also control for a quadratic effect of peoples' (research) age, calculated as the number of years since the occurrence of the first publication in our large data set, starting in 2001. The term δ_i represents individual fixed effects and controls for time-invariant individual attributes. The term γ_t captures year-specific fixed effects that account for changes in publication trends over time. As a consequence of including individual and time fixed effects, the terms $TreatedResearcher_i$ and $AfterKinect_t$ drop out of the estimating equation.

We exploit three categories of dependent variables of interest DV_{it}. First, we estimate changes in researchers' productivity by employing two measures of output: (1) $PubCount_{it}$ captures the number of academic publications of researcher i at time t, and (2) $CitationWeightedPubCount_{it}$ captures the number of academic publications of researcher i at time t weighted by the number of cumulative citations received until 2014. Second, we distinguish between two concepts capturing changes in researches' portfolios of project types: (1) the extent to which the portfolio of researcher i is concentrated or diverse at time t ($Diversification_{it}$) and (2) the extent to which the portfolio of researcher i involves topics that are closer or further from each other in ideas space at time $t-1$ versus t ($Trajectory_{it}$). We do this because a researcher's work could be concentrated in a small number of domains, but these domains could, theoretically, be quite distant from one another. For example, a researcher whose work includes topics in only labor economics and materials science would have a high research concentration, though his or her projects would be quite distant in ideas space. We provide details on the construction of these measures in the next section.

Our main coefficient of interest is β. We interpret a positive value of β as indicating a higher increase in productivity, diversification, or trajectory shift, respectively, at the individual level for treated researchers after the launch of Kinect compared with the change in matched researchers from before to after the launch of Kinect. In other words, a positive value of β indicates a positive effect of the research technology on researchers' rate of knowledge output or their propensity to diversify their research projects or to shift their research trajectories.

We construct a plausible counterfactual using coarsened exact matching (CEM; Iacus and Porro 2011, 2012) based on individual researcher characteristics in the before period (2007–2010). Specifically, we match on (1) yearly productivity in each of the four years before Kinect's launch, (2) the number of coauthors in each of the four years before Kinect's launch, (3) a measure of diversification across knowledge topics between 2007 and 2010, and (4) distance in knowledge space to the motion-sensing domain of knowledge before the launch of Kinect. We measure yearly productivity as a count of publications weighted by citations and diversification as 1 minus the Euclidean distance in the space of IEEE-defined research categories.[8] We define distance in knowledge space to motion-sensing based on the network of authorship with within-area researchers. Specifically, we label within-area researchers as being the closest to the motion-sensing domain of knowledge (distance 1), followed by people who coauthored with within-area researchers on other, non-motion-sensing projects (distance 2), followed by coauthors of coauthors of such researchers (distance 3). All other researchers are categorized as being the furthest away in knowledge space from motion-sensing, to a total of a four-level distance categorization. We conduct two matching procedures, one for each of our two sets of researchers: within-area researchers and outside-area researchers. We include our measure of distance in knowledge space to motion-sensing only in the latter matching procedure, because, by construction, all researchers with an assigned distance of 1 are labeled as treated in the former sample. We employ CEM with weights rather than one-on-one matching to use as much of the available data as possible.

We match on productivity in the before period to ensure that our results on changes in productivity, diversity, and trajectory shift at the individual level are not confounded by researchers at the right tail of the productivity distribution. We match on the number of coauthors in the before period to ensure that our results are not driven by researchers' abilities or preferences for collaborating more intensely or more broadly. This matters for the analysis because higher levels of collaboration could be correlated with more diverse output or with changes in research trajectory because each new collaborator increases the potential pool of expertise and perspectives. We also match on the level of diversification in the pre-Kinect period to ensure that the results are not driven by people with a taste for exploring new avenues that may manifest regardless of research technology availability. Finally, we match on distance to motion-sensing to ensure that our results are not driven by proximity to the motion-sensing field that would influence researchers' project choices regardless of the availability of motion-sensing technology. Our results remain robust to using the full set of researchers in fixed-effects estimations, under the assumption that all researchers publishing in IEEE outlets are at risk for engaging with new technological developments in their research (see online appendix A). Furthermore, raw data trends displayed in Online Appendix A indicate the presence of parallel pretrends ensuring that our CEM procedure does not impose this structure on a phenomenon that follows a different pattern.

4.2. Measuring Changes in the Portfolio of Project Types

We attempt to distinguish between diversification and changes in research trajectories because the two concepts capture research behavior that reflects distinct features of a researcher's portfolio of project types. We define *diversification* as the breadth of a researcher's portfolio of projects at one point in time t. We define *trajectory* as the distance in knowledge space between a researcher's portfolio of projects at times $t - 1$ and t. The measurement challenge is that estimating such changes requires delineating the boundaries of research trajectories. This involves a paradox, however, because the boundaries of research trajectories are part of the core unknown to be estimated. Unlike physical space, which consists of a well-known number of dimensions and distances between locations, ideas space consists of an unknown number of dimensions, the distances between which cannot be uniquely measured, and that evolve over time in ways that cannot be anticipated until they are realized.

Aghion et al. (2008) model the development of ideas along research trajectories. In some cases, such as in mathematics, fields are relatively well defined, and stable field distinctions can form the basis for inquiry about location and movement in ideas space (Borjas and Doran 2012, 2015; Agrawal et al. 2016). In most fields, however, it is difficult to measure such trajectories or identify where they branch. To overcome these challenges, scholars often focus on measures of research breadth (Grupp 1990; Rafols and Meyer 2010) or other measures of topic overlap to reflect whether researchers are roughly in the same domain (Boudreau et al. 2017); consider the development of ideas based on references to papers in a stream of research (e.g., Furman et al. 2012) or rely on a professionally curated tools such as the PubMed related article algorithm (Azoulay et al. 2015, Myers 2018).

These analyses, however, typically rely on observable indicators of innovation output, such as keywords, taxonomies, or citation maps. While helpful in providing some evidence of the evolution of research trajectories, these approaches face many of the challenges described earlier. For example, research that categorizes trajectories based on curated taxonomies has the advantage of consistency but faces the trade-off of either being stable and therefore comparable over time, though at the cost of inflexibility, or being dynamic and therefore evolving with the changing research landscape, though at the cost of consistency and classification standardization. Author-assigned keywords or any other set of keywords not extracted from a defined vocabulary fare better in capturing new knowledge trajectories but lack structure and may be more subject to gaming. This also limits the interpretability of hierarchical connections between keywords and changes in such relationships over time. Similarly, while the citation revolution, as a method for tracing knowledge linkages (Griliches 1990), significantly helped advance our understanding of factors influencing the process of knowledge creation, it is subject to the same concerns because measuring diversity in citation maps requires some form of categorization. As with author-assigned keywords, the selection of backward and forwards citations is subject to social processes and strategic behaviors that complicate their interpretation for understanding changes in research trajectories.[9]

The research context we examine is not characterized by a relatively stable set of keywords and research topics. Indeed, the past two decades have seen the emergence of many new domains of research and associated new keywords enabled by ever-advancing computing power and methods within these fields. To measure research diversification and trajectory in these fields, we leverage advances in machine learning to develop measures that make use of more complete information in academic publications. We propose measures based on topic modeling algorithms, which we have adapted for inference in our context. Our empirical analysis also includes a set of more traditional measures of research diversification and trajectory based on observable characteristics of academic publications, including a measure of diversification based on the stable taxonomy maintained by the IEEE, a measure of research trajectory as a count of new authors, and a measure of research trajectory as a count of new publication outlets.

The main advantage and, hence, contribution of measures based on machine learning analysis[10] is the ability to identify similarities between bodies of text without predefined assumptions about their structure. Note that the intended purpose of these algorithms is prediction, not inference. Their success rests with their ability to reveal the latent structure of a corpus of texts in order to predict with high accuracy where a new text would fit in the structure. We are interested in identifying the latent categorization of research publications in ways that (1) are less subject to the strategic behavior of researchers and (2) are sensitive to the fact that research fields evolve over time.

In the service of these objectives, we employ the hierarchical Dirichlet process (HDP; Teh et al. 2006),[11] which we adapt for our purposes. The algorithm falls into the topic modeling category of unassisted machine learning. HDP is a probabilistic model that employs a hierarchical Bayesian analysis of text (see, e.g., Hofmann 1999, Buntine and Jakulin 2004, Teh et al. 2006). The intuition is that of a generative process in which the data are assumed to be characterized by a set of observed variables (words in the document or vocabulary) that develop from a set of hidden variables (the topic structure) (Teh et al. 2006). The algorithm generates collections of words (topics) that are found to appear together in the input text with a certain probability. In other words, the input text is "assigned" to topics with a certain probability.

We conduct our analysis using the abstracts of the academic publications in our data set as input text into HDP.[12] We run the algorithm per year for the full set of publications available in our data set. We modify the algorithm to output the set of words describing each topic[13] and to list the publication identifications (IDs) of each abstract used to identify those topics. Each publication ID is assigned a score, which can be thought of as a probability of "belonging" to a topic. All scores add up to 100% probability per publication ID.

The algorithm has advantages and limitations. First, HDP has the advantage of identifying the optimal number of topics per corpus of text analyzed. This differs from other topic modeling algorithms, such as latent Dirichlet allocation (LDA), which require the analyst to input the number of topics into which he or she would like the algorithm to group the text. We also run a sensitivity algorithm on multiple instances of LDA, consistent with state-of-the-art practices in computer science, to identify the number of optimal topics with the highest probability of accuracy.[14] Second, both algorithms treat the input text as a one-time group for which the latent categorization needs to be revealed. In other words, the algorithms cannot automatically track the evolution of topics over time by updating the set of keywords in each category over time. We address this limitation by calculating a cosine vector similarity between the yearly topics generated by the HDP algorithm. Specifically, we employ a term frequency–inverse document frequency (TFIDF) cosine similarity where the frequency of words is weighted by the HDP-generated score that captures the relevance of each word for each topic. In addition, we use the HDP output in a regression with time fixed effects; hence, our results are not hindered by the fact that the algorithms are executed on a per-year basis and thus reveal the latent categorization of topics for each year in our data set.[15]

We calculate diversification as a yearly measure of spread across categories, as identified by the HDP algorithm. Specifically, we calculate an intensive measure of diversification equal to the sum of the number of topics where the focal researcher has his or her papers assigned by the HDP in the focal year. We also calculate an extensive measure of diversification equal to a count of unique topics where the focal researcher has his or her papers assigned by the HDP in the focal year. We also present results using more traditional measures of diversification based on publication attributes. Specifically, we calculate a yearly reversed Euclidean distance in the space of 51 IEEE categories in computer science, electrical engineering, and electronics. To calculate this measure, we apply Equation (2) to yearly publication data over the period of interest (2007–2014), four years before and four years after the launch of Kinect.

To generate measures that capture changes in research trajectory, we use the yearly topics generated by the HDP and the cosine similarity index between such topics. Specifically, we first calculate the distance between topics in consecutive years as one minus the similarity index between all such topic pairs. Next, for each researcher, we sum the distance between all pairs of topics in year $t - 1$ and t and then divide the sum by the number of unique topics covered in year $t - 1$. The measure captures the average number of new topics for researcher i in year t relative to year $t - 1$, weighted by the distance between the topics. One can think of this measure as capturing the yearly new areas of interest; the smaller the value, the less of a change in interests there will be from one year to the next.

In addition, we employ two other measures based on traditional publication output. The first such measure counts the number of new coauthors that the focal author has in the observation year relative to previous years. To count the number of new coauthors, we take advantage of our full data set going back to 2001. We do so because, by definition, the count of new coauthors requires a few years of reference data to get closer to reflecting the true number of new coauthors and to not be upward biased owing to left-side data truncation. This measure indicates changes in research trajectories to the extent that collaboration patterns reflect changes in the bases of expertise associated with a researcher's project choices. The second measure reflects the number of new publication outlets in which a researcher publishes each year, relative to each prior year, going back to 2001. This measure indicates changes in research trajectories to the extent that different journals address different audiences and cover different areas of the ideas space.

Each measure has its own limitations and merits. The HDP-based measures are more flexible in capturing changes in knowledge space over time. However, the HDP categorization lacks stability and trackability over time. The IEEE diversification measure fairs well on these dimensions, but its disadvantage stems from the same attributes that we discussed earlier. In particular, the fixed taxonomy fails to capture changes in the categorization structure over time that would otherwise indicate changes in research trajectories. Similarly, the measures based on counts of new coauthors and new publication outlets, while perhaps easier to grasp than interpreting the HDP measure, are focused on outcomes that indirectly reflect the content or intellectual focus of academic research. Specifically, it is possible to change coauthors and publication outlets while continuing to work on the same knowledge trajectory, and it is also possible to continue working with old coauthors and publish in the same journals while shifting one's research focus.

As a result, we prefer our more novel measures, which we think are more likely to reflect changes in research diversification and trajectory because they are more tied to the content of researchers' published works. At the same time, we believe that the other measures complement the HDP measures and enrich the insights than can be drawn from our analysis. Each measure captures different attributes of diversification and trajectory. Taken together, we argue, they paint an informative picture of changes in researchers' projects triggered by the availability of automating research technology.

3.3. Sample Construction and Descriptive Statistics

We collect data on every publication, early-access publication, and conference proceeding paper available through IEEE Xplore between 2001 and 2014. These data include 2,492,451 publications and 1,670,888 unique author names in the fields of computer science, electrical engineering, and electronics. Because of the importance of publications in conference proceedings in computer science and related fields, the IEEE database possesses advantages relative to other libraries of publications, including the Web of Science and Scopus.

We focus our analysis on the four years leading up to and the four years following the launch of Kinect at the end of 2010, that is, 2007–2010 and 2011–2014. We do so to ensure comparable timeframes and to allow for some publication data for controls and other measures that require a longer-run observation of publication trends, such as author age and changes in the number of yearly new coauthors and new publication outlets. The 2007–2014 data set consists of 1,776,125 publications authored by 1,391,313 people as identified by the IEEE. Within this subset, we further distinguish between researchers active both before and after Kinect's launch (430,779), only in the period before 2010 (442,395), and researchers who enter the sample after 2010 (518,139). We do so (1) to ensure that our main results are not driven by zeroes as a result of exits from or entry into our observation period and (2) to allow for an observable period before Kinect's launch from which we can identify trends and construct plausible counterfactuals. We focus our main analysis on the first data set, and we further eliminate outliers, namely people with fewer than three and more than 50 publications before Kinect's launch (2007–2010). We eliminate researchers with fewer than three publications because some of our measures rely on peoples' breadth of publications, and low productivity mechanically translates into low diversification. Note that this set of researchers also includes authors who occasionally publish in outlets tracked by the IEEE. We eliminate authors with more than 50 publications in the before period to account for potential disambiguation effects in the IEEE algorithm assigning unique author identifiers. Such author IDs would appear as very productive and potentially diversified individuals, risking an upward bias to our diversification-based estimations. We identify a total of 3,200 researchers with over 50 publications in the four-year period before Kinect's launch, less than 1% of our main sample.[16]

Our final sample includes 12,549 within-area researchers, 9,590 outside-area researchers, and 160,845 other researchers. We present descriptive statistics that show the balance in our CEM procedure for within-area and outside-area researchers in Tables B1.a and B1.b of Online Appendix B. We show average values for all our covariates used in the matching procedure, in both the full sample (columns (1)–(3)) and the matched sample (columns (4)–(6)). Across all definitions, treated researchers are more productive.[17] They have a higher number of coauthors and publish in more IEEE categories than the rest of the population. However, this fact results from a more skewed distribution in the full sample. The CEM procedure balances these observables for each of our two definitions of treated researchers. Our sample size is reduced by the matching procedure and by our eliminating outliers, as described earlier.

In Tables 1–3, we present average values of all our dependent variable measures across the two groups of treated researchers: two measures of publication rate, three measures of diversification, and three measures of research trajectory. The descriptive statistics foreshadow our main findings: (1) an increase in quality publication output for both within- and outside-area researchers, (2) a positive impact on diversification that is more pronounced for outside-area researchers than for within-area researchers, and (3) a positive impact on the trajectory of research that is more pronounced for outside-area researchers than for within-area researchers. Further, we observe differences across all our measures relative to baseline trends. Some of the effects come from mitigating the overall decrease in a productivity, diversification, or trajectory shift, whereas others come from an accentuated baseline trend of an increasing productivity, diversification, or trajectory shift, respectively. Specifically, the increase in productivity measured by the count of academic publications is absolute and offsets an overall trend of a decrease in publication output. Our measure of productivity based on publication output weighted by citation count offsets an overall downward-sloping trend that can be explained not only by the average decrease in academic publication but also by the skewed nature of citation accumulation over time. Similarly, our Euclidean measure of

Table 1. Changes in Researchers' Productivity for Our Matched Sample: Descriptive Statistics

Descriptive statistics: Rate (mean (standard deviation (SD)))

Treated as		Publication count		Publication count weighted by citations	
		Treated	Controls	Treated	Controls
Within-area researchers	Before	1.182 (1.241)	1.214 (1.295)	2.291 (3.346)	2.256 (3.327)
	After	1.249 (1.669)	1.153 (1.600)	1.767 (3.034)	1.585 (2.941)
Observations		13,168	111,936	13,168	111,936
Outside-area researchers	Before	1.068 (1.199)	1.086 (1.240)	1.900 (3.048)	1.891 (3.058)
	After	1.676 (1.994)	1.022 (1.469)	2.408 (3.740)	1.404 (2.672)
Observations	Before	11,936	79,008	11,936	79,008

diversification, which captures the information contained in IEEE's taxonomy relative to changes in the ideas space, shows an overall trend of decreasing diversification, with the research technology dampening that effect. The same trends can be observed in our measure of diversification based on counts of new publication outlets. This is important for at least two reasons. First, it underscores our point regarding the difficulty of crafting all-encompassing measures of diversification and research trajectory. Second, it highlights the differences in the various aspects of the knowledge creation process captured by each measure. For example, the overall trend of a decrease in diversification as captured by the Euclidean measure appears to contrast with the overall trend of an increase in diversification captured by our topic modeling measures. This apparent inconsistency may, however, result from an increase in the fragmentation of research trajectories, an effect aligned with the "knowledge burden" effect (Jones 2009, 2010) that cannot be captured by a measure that relies on a fixed taxonomy. More broadly, the differences suggest a need for approaching studies on changes in research diversification and trajectories using empirical strategies that triangulate across multiple measures.

5. Empirical Analysis

5.1. Did the Kinect Shock Induce Changes in Researchers' Productivity?

We begin our analysis by estimating the impact of Kinect on the researchers' productivity. To do this, we estimate Equation (1) using the annual count of

Table 2. Changes in Researchers' Diversification for Our Matched Sample: Descriptive Statistics

Descriptive statistics: Diversification (mean (SD))

Treated as		HDP topic count, intensive		HDP topic count, extensive		Euclidean diversification using IEEE taxonomy	
		Treated	Controls	Treated	Controls	Treated	Controls
Within-area researchers	Before	4.106 (4.634)	4.098 (4.654)	3.091 (2.951)	3.017 (2.890)	0.360 (0.296)	0.358 (0.294)
	After	4.899 (6.858)	4.455 (6.489)	3.189 (3.356)	2.919 (3.216)	0.313 (0.288)	0.283 (0.285)
Observations		13,168	111,936	13,168	111,936	13,168	111,936
Outside-area researchers	Before	3.642 (4.432)	3.662 (4.537)	2.751 (2.872)	2.709 (2.867)	0.325 (0.296)	0.323 (0.297)
	After	6.530 (8.198)	3.934 (5.991)	3.913 (3.583)	2.665 (3.102)	0.372 (0.286)	0.262 (0.281)
Observations		11,936	79,008	11,936	79,008	11,936	79,008

Table 3. Changes in Researchers' Trajectory for Our Matched Sample: Descriptive Statistics

Descriptive statistics: Trajectory (mean (SD))

Treated as		Yearly new HDP topics by distance		Yearly new coauthors		Year new publication outlets	
		Treated	Controls	Treated	Controls	Treated	Controls
Within-area researchers	Before	3.466 (5.599)	3.114 (5.163)	2.242 (3.086)	2.258 (3.074)	0.839 (0.964)	0.883 (1.014)
	After	4.415 (7.133)	3.970 (6.525)	2.565 (4.555)	2.359 (4.387)	0.789 (1.111)	0.739 (1.055)
Observations		13,168	111,936	13,168	111,936	13,168	111,936
Outside-area researchers	Before	2.957 (5.156)	2.707 (5.085)	2.061 (3.428)	2.048 (3.367)	0.783 (0.951)	0.806 (0.973)
	After	5.613 (8.208)	3.554 (6.105)	3.512 (5.574)	2.126 (4.195)	1.063 (1.298)	0.674 (0.991)
Observations		11,936	79,008	11,936	79,008	11,936	79,008

Table 4. Estimated Changes in Publication Count

Controls determined through CEM

	Count of publications		
	Within-area researchers	Outside-area researchers	Outside-area researchers (exclude motion-sensing papers)
Treated × *AfterKinect*	0.083*** (0.017)	0.497*** (0.017)	0.305*** (0.019)
Quadratic age	Yes	Yes	Yes
Individual and year fixed effects (FEs)	Yes	Yes	Yes
Log likelihood	−270,358.60	−187,776.56	−186,346.70
Observations	231,666	166,212	166,212

Notes. The data are a panel at the researcher level based on publication data between 2007 and 2014. All models are Poisson with robust standard errors clustered at the individual level. CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

academic publications at the individual level as the dependent variable as well as an annual count weighted by citations accumulated by 2014. We should note that we observe this effect in descriptive data, because the average number of publications for within-area researchers increases from 1.18 to 1.25, whereas in the matched control group the average number of publications changes from 1.21 to 1.15. In the group of outside-area researchers, the average number of publications changes from 1.07 to 1.68, whereas in the control group it changes from 1.09 to 1.02.

We show the results of our main difference-in-differences estimation in Table 4. In column (1), we estimate the effect on the productivity of within-area researchers. We find evidence of a 9% increase in publication count for within-area researchers compared with matched controls, or one additional paper for every 10, in line with Hypothesis 1a. Column (2) shows results for outside-area researchers, and column (3) shows results for the same group of outside-area researchers when excluding their motion-sensing publications. In both cases, we find evidence of an increase in the productivity of such researchers compared with the group of control researchers, as identified through our CEM procedures, in line with Hypothesis 2a. Treated outside-area researchers increase their number of publications by 65% more than the control group, an effect that corresponds to an increase of three additional papers for every two.[18] In column (3), we show evidence that the increase in publication output is not driven by publications in motion-sensing. This is important because it suggests that the effect of research technology does not lead mechanically to increases in productivity via direct engagement with the technology, but rather access to the research technology affects the productivity of outside-area researchers on all topics on which they work, in line with our Hypothesis 2a. Specifically, outside-area researchers experience a disproportionate increase of 36%, approximately one additional paper for every three publications, in their publications on topics outside of motion-sensing. Furthermore, the magnitude of the effects is higher for outside-area researchers than for within-area researchers,[19] in line with Hypothesis 2a. All these trends persist when focusing on a measure of publication output weighted by citations (Table 5), suggesting that the increase in productivity does not arise as a result of a jump in studies of lesser quality.

Next, we examine the timing of this effect in order to ensure that our estimates of the boost in

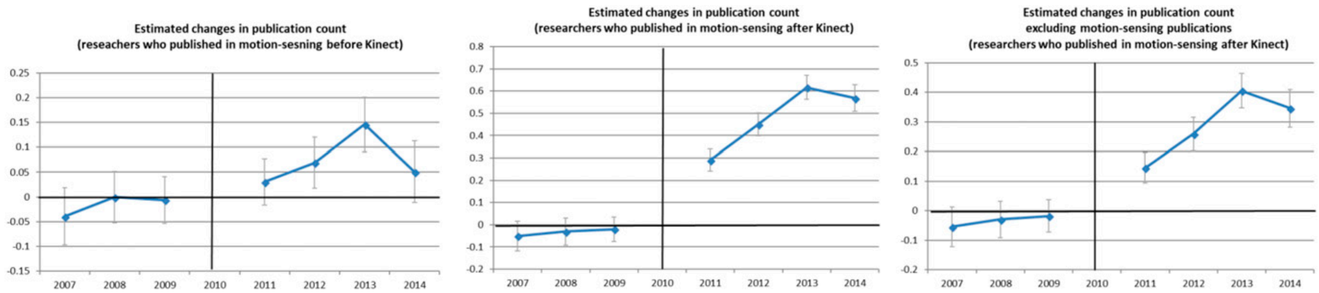
Table 5. Estimated Changes in Citation Weighted Publication Count

Controls determined through CEM

	Count of publications weighted by citations		
	Within-area researchers	Outside-area researchers	Outside-area researchers (exclude motion-sensing papers)
Treated × *AfterKinect*	0.083*** (0.023)	0.537*** (0.024)	0.330*** (0.027)
Quadratic age	Yes	Yes	Yes
Individual and year FEs	Yes	Yes	Yes
LL	−424,738.03	−285,514.28	−283,379.44
Observations	231,666	166,212	166,212

Notes. The data are a panel at the researcher level based on publication data between 2007 and 2014. All models are Poisson with robust standard errors clustered at the individual level. CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 1. Estimated Changes in Yearly Publication Counts



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in publication counts between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

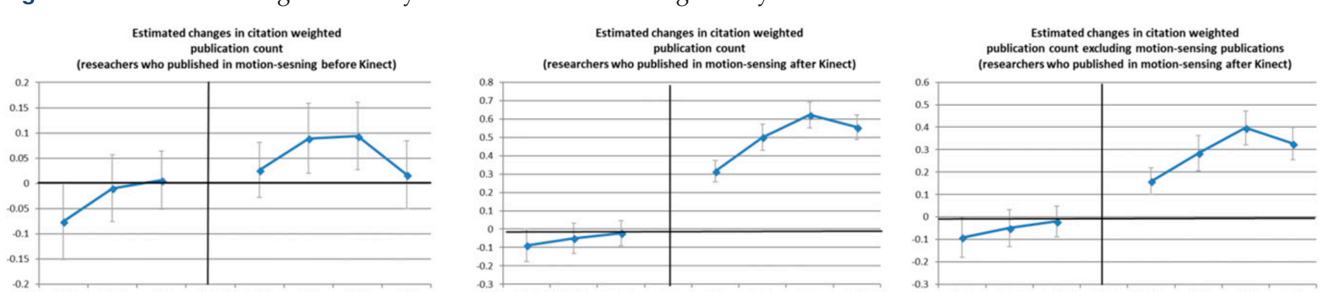
productivity are not driven by secular trends toward increased publication or changes in productivity that occur later during our after period and which could therefore be attributed to events other than the launch of Kinect. To do this, we replace $AfterKinect_t$ with a set of year-specific dummy variables. We plot the estimated value of the interaction between these year dummies with our treated dummy in Figures 1 and 2. All estimates consider 2010 as our baseline year (Kinect was launched on November 4, 2010). Each point on the graphs represents the estimated difference between the number of publications of treated versus control researchers, for each year, relative to the same difference in 2010. In each graph, the difference is small and close to zero before the launch of Kinect, as expected given our CEM procedure, and increases immediately thereafter, consistent with the conclusion that the availability of motion-sensing technology triggered an increase in researchers' productivity. In line with the discussion of coefficient magnitudes in Tables 4–5, the effect is higher for outside-area researchers than for within-area researchers.

5.2. Did the Kinect Shock Induce Research Diversification?

Our first evidence comes from descriptive statistics on entry into motion-sensing research. Our estimation strategy precludes including authors who published only in the post-Kinect period; that is, we do not observe a before period in order to identify an appropriate counterfactual. In Table 6, we report the number of new unique authors (based on IEEE database author identifiers) that appear in the data set over the eight-year course of our sample. We distinguish between researchers entering our sample with at least one motion-sensing publication in their first year (column (1)) and researchers entering with other types of publications (column (2)). Column (3) reports the ratio of new entries in motion-sensing research. These data suggest diversification via entry into motion-sensing research. Prior to the Kinect launch, approximately 1% of entries into the IEEE data set occurs via publications in motion-sensing, but this percentage increases by 31% immediately following the Kinect launch and continues to increase, up to 2.5 times in 2014, relative to the period before Kinect.

To investigate this question in a more formal way, we return to our main difference-in-differences estimation from Equation (1). First, we focus on the impact on within-area researchers in Tables 7 and 8. Table 7 presents estimates using a dummy variable for treated people, and Table 8 shows estimates using a continuous measure of involvement in motion-sensing before Kinect. Specifically, we replace $TreatedResearcher_i$ in Equation (1) with a variable capturing the level of specialization in motion-sensing before Kinect, calculated

Figure 2. Estimated Changes in Yearly Publication Counts Weighted by Citations



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in publication counts between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

Table 6. New Researcher Entries in Electrical Engineering, Computer Science, and Electronics, per Year, as Observed through Publications Logged in the IEEE Xplore Database, 2007–2014

	Number of authors entering IEEE Xplore academic publication		
Year	Number of authors entering with at least one motion-sensing publication	Number of authors entering with publications in other areas	Percentage of motion-sensing entry
2007	1,103	122,322	0.90
2008	1,307	124,670	1.04
2009	1,188	120,554	0.99
2010	1,502	152,237	0.99
2011	1,794	137,233	1.31
2012	2,449	135,514	1.81
2013	2,805	116,184	2.41
2014	2,985	119,265	2.50

Table 7. Estimated Changes in the Level of Diversification of Within-Area Researchers

Treated as within-area researchers; controls determined through CEM			
Diversification	HDP topic count (intensive)	HDP topic count (extensive)	Euclidean diversification using IEEE taxonomy
Treated × *AfterKinect*	0.075*** (0.018)	0.052*** (0.013)	0.025*** (0.004)
Quadratic age	Yes	Yes	Yes
Individual and year FEs	Yes	Yes	Yes
LL/R^2	−765,976.57	−527,561.97	0.045
Observations	231,263	231,263	231,263

Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Columns (1) and (2) estimate a Poisson model. Column (3) estimates a linear regression model (ordinary least squares (OLS)). CEM, coarsened exact matching.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

as the sum of motion-sensing publications over the period before the launch of Kinect (2007–2010) and divided by the total publication output of researcher i over the same period. The table presents estimates using our three different diversification measures, the intensive and extensive HDP topic counts and the Euclidean diversification measure based on the IEEE taxonomy.

Table 7, columns (1) and (2), shows evidence of an 8% increase in the number of HDP topics covered yearly and a 5% increase in the number of unique HDP topics covered yearly relative to the control group, consistent with Hypothesis 1b. This amounts to a modest increase of 0.33 topic (and 0.16, respectively) of approximately 50 topics covering the complete set of research in computer science and engineering. These findings are consistent with those obtained using the traditional diversification measure (which indicates a 7% increase in the Euclidean-based diversification) but have the added benefit of ease of interpretation of magnitude relative to the number of additional topics. Table 8 shows that the increase in diversification is more pronounced for those who had previously focused their research in

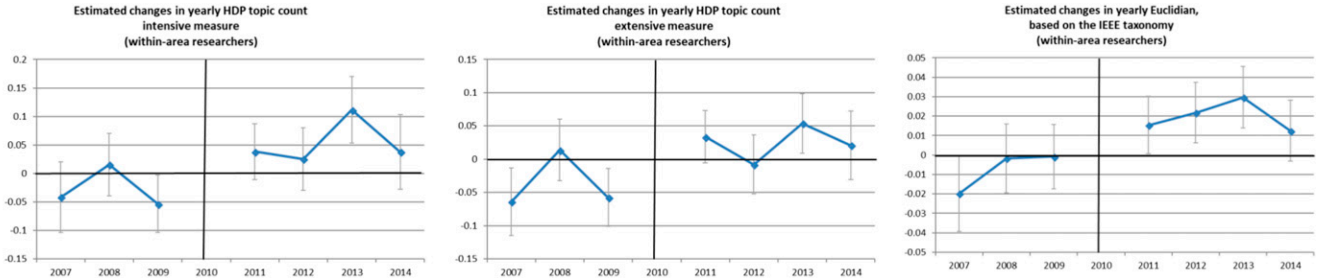
Table 8. Specialists in Motion-Sensing Experience a Higher Level of Change in Diversification after Kinect

Treated as within-area researchers; controls determined through CEM			
Diversification	HDP topic count (intensive)	HDP topic count (extensive)	Euclidean diversification using IEEE taxonomy
Fraction of MS publications before × *AfterKinect*	0.195*** (0.042)	0.146*** (0.030)	0.064*** (0.009)
Quadratic age	Yes	Yes	Yes
Individual and year FEs	Yes	Yes	Yes
LL/R^2	−765,971.96	−527,554.55	0.045
Observations	231,263	231,263	231,263

Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Columns (1) and (2) estimate a Poisson model. Column (3) estimates a linear regression model (ordinary least squares (OLS)). CEM, coarsened exact matching.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 3. Estimated Changes in the Level of Diversification of Within-Area Researchers



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

motion-sensing as opposed to researchers who also published in other domains. This suggests that rather than frustrating opportunities for within-area researchers, the automating technology facilitates exploration of ideas in their domains of expertise.

As before, we test the timing of these effects by replacing the postshock dummy, $AfterKinect_t$, with a series of dummy variables reflecting each year of the analysis. We plot the estimated difference in yearly level of diversification between our treated and control researchers in Figure 3. All values are computed relative to 2010. In each case, we confirm the absence of pretrends as constructed through our CEM procedure and find evidence of an increase in average researcher diversification that begins in the year after the Kinect launch and persists across time.

Next, we turn our attention to the effect for outside-area researchers and show results in Tables 9 and 10. As before, Table 9 presents estimates using our three different diversification measures, the intensive and extensive HDP topic counts and the Euclidean diversification measure based on the IEEE taxonomy, considering the motion-sensing publications produced by these researchers after the launch of Kinect. Table 10

Table 9. Estimated Changes in the Level of Diversification of Outside-Area Researchers

Treated as outside-area researchers; controls determined through CEM			
Diversification	HDP topic count (intensive)	HDP topic count (extensive)	Euclidean diversification using IEEE taxonomy
Treated × *AfterKinect*	0.500*** (0.019)	0.358*** (0.013)	0.103*** (0.004)
Quadratic age	Yes	Yes	Yes
Individual and year FEs	Yes	Yes	Yes
LL/R^2	−537,282.06	−373,937.94	0.048
Observations	165,868	165,868	166,212

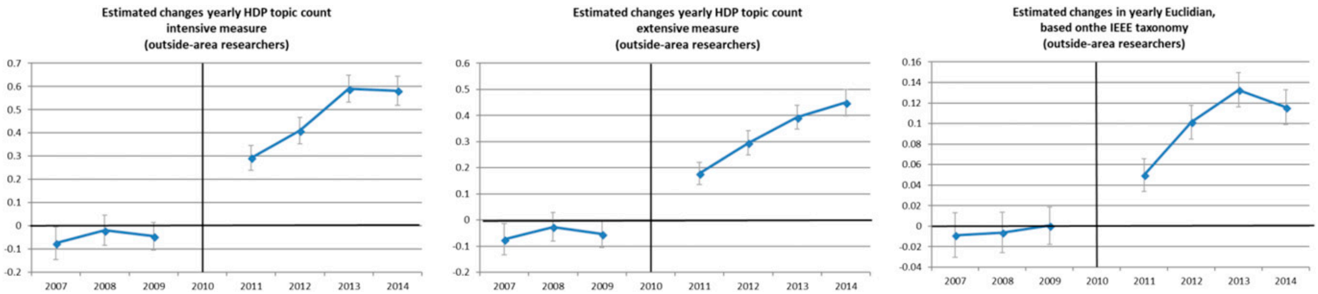
Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Columns (1) and (2) estimate a Poisson model. Column (3) estimates a linear regression model (OLS). CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 10. The Change in Diversification of Outside-Area Researchers Persists Outside the Set of Newly Added Motion-Sensing Publications

Treated as outside-area researchers; controls determined through CEM			
Diversification	HDP topic count (intensive)	HDP topic count (extensive)	Euclidean diversification using IEEE taxonomy
Treated × *AfterKinect*	0.310*** (0.021)	0.191*** (0.015)	0.048*** (0.005)
Quadratic age	Yes	Yes	Yes
Individual and year Fes	Yes	Yes	Yes
LL/R^2	−534,765.16	−372,990.12	0.046
Observations	165,844	165,844	166,212

Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Columns (1) and (2) estimate a Poisson model. Column (3) estimates a linear regression model (OLS). CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 4. Estimated Changes in the Level of Diversification of Outside-Area Researchers



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

eliminates the motion-sensing publications and focuses on the impact on diversification for all other projects of these researchers.

Table 9, columns (1) and (2), shows evidence of a 65% increase in the number of HDP topics covered yearly and a 43% increase in the number of unique HDP topics covered yearly, relative to the control group, in line with Hypothesis 2b. This amounts to an increase of 2.37 topics (and 1.18, respectively) of approximately 50 topics. These findings are consistent with those obtained using the traditional diversification measure, which indicates a 32% increase in Euclidean-based diversification, but with the added benefit of ease of interpretation of magnitude relative to the number of additional topics. Additionally, we observe that the effect on outside-area researchers is larger than that on within-area researchers, in line with Hypothesis 2b. As before, we include more formal estimates of the difference in magnitude in Table C2 of Online Appendix C.

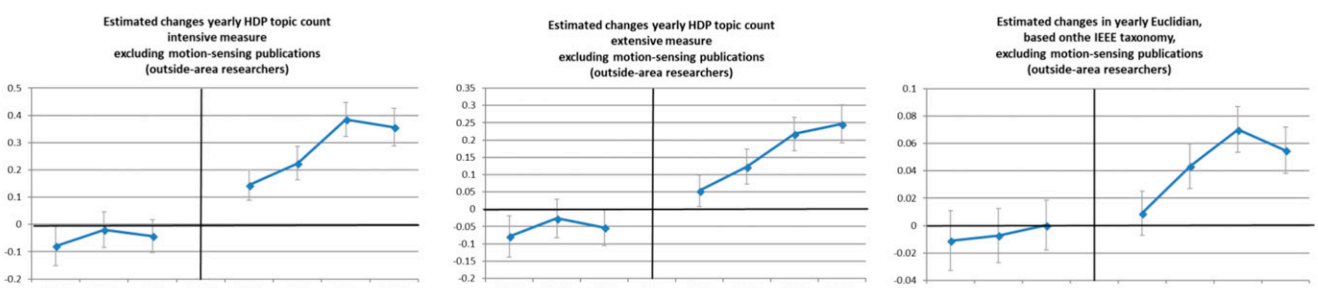
Table 10 demonstrates that this boost in diversification persists when we exclude motion-sensing publications. Specifically, the automation of motion-sensing technology facilitates a 36% increase in the number of HDP topics covered yearly by all other non-motion-sensing publications produced by these researchers relative to the matched control group. Similarly, column (2) indicates a 21% increase in the number of unique HDP topics covered yearly by all other non-motion-sensing publications produced by these researchers, relative to the control group. The change in diversification amounts to an increase of 1.31 topics (and 0.58, respectively).

As before, we test the timing of these effects by replacing $AfterKinect_t$ with a set of dummy variables for each sample year. We plot the estimated difference in yearly level of diversification between our treated and control researchers in Figures 4 and 5. All values are computed relative to 2010. In each case, we confirm the absence of pretrends as constructed through our CEM procedure and find evidence of an increase in average researcher diversification following the launch of Kinect that begins immediately after the launch and persists across time.

5.3. Did the Kinect Shock Induce Changes in Researchers' Trajectories?

To turn to the question of whether the Kinect shock induced changes in researchers' trajectory of inquiry, we apply the same estimation strategy as in the preceding section but replace the diversification variables with three alternative measures designed to

Figure 5. Estimated Changes in the Level of Diversification of Outside-Area Researchers, When Excluding Their Motion-Sensing Publications



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

Table 11. Estimated changes in the trajectory of research of within-area researchers

Treated as within-area researchers; controls determined through CEM			
Trajectory	Yearly new HDP topics by distance	Yearly new coauthors	Year new publication outlets
Treated × AfterKinect	0.241** (0.096)	0.051** (0.020)	0.074*** (0.017)
Quadratic age	Yes	Yes	Yes
Individual and year Fes	Yes	Yes	Yes
R^2/LL	0.028	–527,103.78	–208,371.97
Observations	231,266	230,770	231,026

Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Column (1) estimates a linear regression model (OLS). Columns (2) and (3) estimate a Poisson model. CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

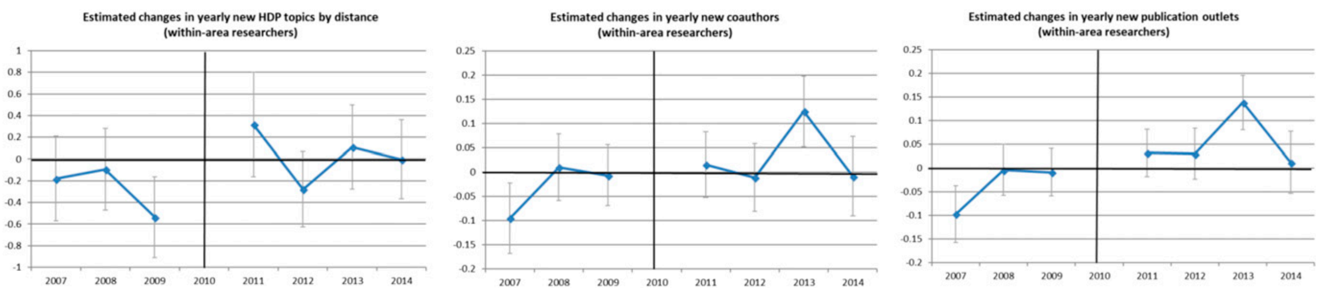
Table 12. Specialists in Motion-Sensing Experience a Higher Level of Shift in Their Trajectory of Research

Treated as within-area researchers; controls determined through CEM			
Trajectory	Yearly new HDP topics by distance	Yearly new coauthors	Year new publication outlets
Fraction of MS publications before × AfterKinect	0.859*** (0.196)	0.170*** (0.051)	0.228*** (0.039)
Quadratic age	Yes	Yes	Yes
Individual and year Fes	Yes	Yes	Yes
R^2/LL	0.028	–527,090.78	–208,366.32
Observations	231,266	230,770	231,026

Notes. The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Column (1) estimates a linear regression model (OLS). Columns (2) and (3) estimate a Poisson model. CEM, coarsened exact matching.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

reflect differences in the distance in knowledge space between researchers' portfolios of projects at times $t-1$ and t. Tables 11 and 12 present the results of changes in trajectory for within-area researchers. Tables 11 presents estimates using a dummy variable for treated people, and Table 12 shows estimates using a continuous measure of involvement in motion-sensing before Kinect. Column (1) of Table 11 estimates that within-area researchers experience a 7% increase in new topics, equivalent to a modest shift of 0.24 topic. Column (2) estimates a 5% increase in the number of new coauthors (1 in 10 new coauthors), and column (3) estimates an 8% increase in the number of new publication outlets (1 in 15 new publication outlets). The advantage of the HDP measure is that it allows us to estimate the number of additional topics the researcher undertakes; an increase in the number of authors and publication outlets does not necessarily imply an increase in the number of topics because researchers can work with new coauthors and publish in different journals without changing their topics of interest or, alternatively, by abandoning old topics and undertaking new ones. The effects are most pronounced for researchers who have a greater degree of involvement in motion-sensing research before Kinect (Table 12). We test the timing of these effects and display the yearly estimated coefficients in Figure 6. We interpret the results as implying that

Figure 6. Estimated Changes in the Type of Knowledge Output of Within-Area Researchers



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher × Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

Table 13. Estimated Changes in the Trajectory of Research of Outside-Area Researchers

Treated as outside-area researchers; controls determined through CEM			
Trajectory	Yearly new HDP topics by distance	Yearly new coauthors	Year new publication outlets
Treated × AfterKinect	1.962*** (0.110)	0.461*** (0.022)	0.448*** (0.016)
Quadratic age	Yes	Yes	Yes
Individual and year Fes	Yes	Yes	Yes
R^2/LL	0.035	−372,894.72	−145,186.49
Observations	166,212	165,547	165,876

Notes: The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Column (1) estimates a linear regression model (OLS). Columns (2) and (3) estimate a Poisson model. CEM, coarsened exact matching.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Kinect induced within-area researchers to expand their research into areas in which they had not worked before, in line with our hypothesis Hypothesis 1b.

Tables 13 and 14 suggest that the same holds—though to a larger degree[20]—for outside-area researchers, in line with Hypothesis 2b. As in prior analyses, Table 13 presents estimates of the three trajectory measures, including motion-sensing publications produced by these researchers after the launch of Kinect. Table 14 eliminates the motion-sensing publications and focuses on the impact on trajectory for all other projects of these researchers. Table 13, column (1), indicates a 66% increase in new topics, equivalent to a shift in trajectory of two new topics. Column (2) of the same table estimates a 59% increase in the number of new coauthors (1.25 new coauthors), and column (3) estimates a 57% increase in the number of new publication outlets (1 in 2 additional publication outlets). The effects persist when eliminating the set of motion-sensing publications from the portfolio of these researchers, suggesting that the effect is not mechanically driven by incorporating motion-sensing into the portfolio of projects (Table 14), in line with Hypothesis 2b. Specifically, outside-area researchers increase the number of yearly new topics by 22% (0.45 new topic of approximately 50), the number of new coauthors by 30% (more than 1 in 2 new coauthors), and the number of new publication outlets by 61% (1 in 2 new outlets). As before, we test the timing of these effects and display the yearly estimated coefficients in Figures 7 and 8.

5.4. Complementary Descriptive Analyses

The results of our regression analyses suggest that the availability of this motion-sensing technology induces greater research diversity and a shift in the trajectory of researchers' project portfolios. Ideally, we also want to gain insight into the type of knowledge these people create as part of the identified effects. However, the difficulty of capturing changes in the type of knowledge produced also poses challenges to our ability to speak directly to the measured increase in diversification and shift in trajectory. To shed some light on this, in addition to the examples mentioned earlier, we provide some descriptive examples we collected through a different bibliographic database, Scopus, and its analyze function, which allows user to select a group of papers and analyze their attributes in terms of, for example, spread across domains of knowledge, counts of authors affiliated with various institutions, and spread across different publication outlets. We observe two broad avenues through which researchers enhance the trajectory of their

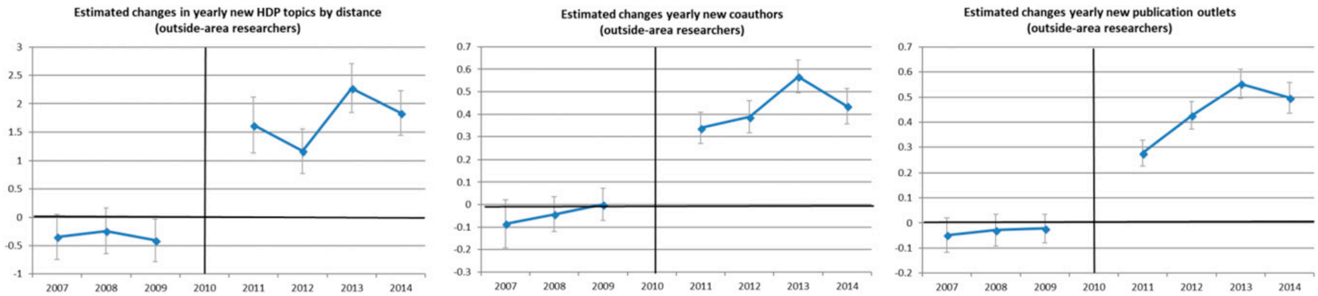
Table 14. The Shift in the Trajectory of Outside-Area Researchers Persists Outside the Set of Newly Added Motion-Sensing Publications

Treated as outside-area researchers; controls determined through CEM			
Trajectory	Yearly new HDP topics by distance	Yearly new coauthors	Year new publication outlets
Treated × AfterKinect	0.451*** (0.052)	0.261*** (0.024)	0.474*** (0.016)
Quadratic age	Yes	Yes	Yes
Individual and year Fes	Yes	Yes	Yes
R^2/LL	0.028	−370,110.33	−145,752.72
Observations	166,212	165,515	165,876

Notes: The data are a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. Column (1) estimates a linear regression model (OLS). Columns (2) and (3) estimate a Poisson model. CEM, coarsened exact matching.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Figure 7. Estimated Changes in the Type of Knowledge Output of Outside-Area Researchers

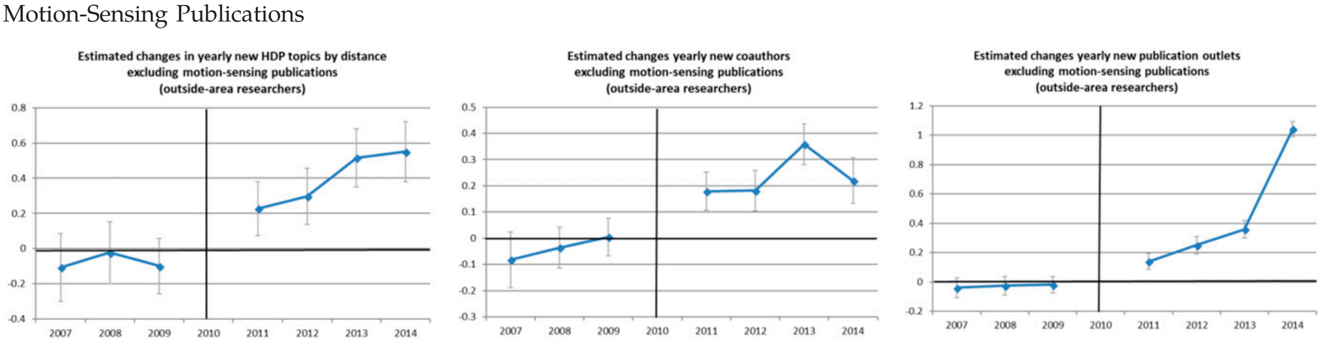


Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in our direction measures between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

inquiry following Kinect: (1) an increase in the number of areas where they publish and (2) an increase in the percentage of their publications across knowledge areas in their portfolio. We include these data and additional information in Online Appendix D.

Additionally, we use the Scopus analyze feature to provide a glimpse into the mechanism through which the benefits of the automating technology manifest by attempting to identify the types of institutions that benefit most from the availability of Kinect as automating motion-sensing research technology.[21] On the one hand, the benefit of Kinect as a technology that automates research tasks might manifest more for research institutions that are or aim to be leaders in research. The assumption is that these institutions are time constrained rather than financially constrained, and the automation would help propel their productivity forward. On the other hand, the benefit of Kinect as technology that reduces the costs of performing certain research tasks might manifest more for institutions that are financially constrained. It follows that a research technology that is automating by significantly reducing the cost of performing certain research tasks should benefit both types of institutions. It is important to remember that we cannot separate these effects from those resulting from the low monetary cost of Kinect. In addition to reducing the cost of executing certain research tasks via automation, Kinect's low monetary cost (relative to the previous generation of motion-sensing technology) may have enabled experimentation that might not have been considered if Kinect were priced in the same range as the previous generation technology. It may be, therefore, that changes in the monetary cost of motion-sensing research technology drive some of the effects we observe. However, if these effects were dominant, then we would expect to see financially constrained institutions disproportionally benefiting from the Kinect phenomenon. In fact, when collecting author affiliations for the top 25% and top 10% most cited motion-sensing publications during the period before and after Kinect, we observe that the set of institutions experiencing the highest increase in the number of top-cited publications is comprised of both highly ranked universities and private institutions with few financial constraints and lower-ranked institutions with presumably more financial constraints. This finding is aligned with the idea that the automating research technology reduces the cost of performing certain research tasks by

Figure 8. Estimated Changes in the Type of Knowledge Output of Outside-Area Researchers, When Excluding Their Motion-Sensing Publications



Notes. We base this figure on our 2007–2014 data set. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher* × *Year* and thus describes the relative difference in our direction measures between treated and control authors in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 2010.

substituting for human capital and thus benefiting both leading institutions racing to maintain their frontier position and more financially constraining institutions. We include these data and additional information in Online Appendix E.

6. Discussion and Conclusion

Automating technologies are transforming the nature of work and competition across industries. Scholarship in management and economics has investigated the impact of such technologies on the manufacturing and service sectors and has documented the extent to which and the conditions under which these technologies substitute for or complement human labor. Although research on innovation emphasizes the importance of research tools, less work investigates the role of such automating technologies in the production of knowledge. In this paper, we contribute to addressing this gap by examining the impact of an automating technology on the rate and type of knowledge production.

Ideas production differs in key ways from the production of goods and services. We believe that a core contribution of this paper involves training a lens on the way in which automating technologies affect the organization of knowledge production. In contrast to some traditional manufacturing positions (e.g., assembly-line work) and service positions (e.g., fast-food preparation), knowledge work is characterized by a particularly wide range of tasks (e.g., problem selection, grant writing, data analysis, paper composition, and seminar presentation) and a substantial amount of autonomy for knowledge workers to select their tasks. This task variety and authority over their own work may enable knowledge workers to adapt to the introduction of automating technologies in ways that vary from those of workers—and even managers—in manufacturing and service sectors. Our results suggest that this is the case. Researchers with prior involvement in motion-sensing are not negatively impacted by the availability of motion-sensing technology that automates core research tasks. Moreover, such researchers, on average, accelerate the pace of their work in response to automation and spread their knowledge wider.

A second area to which we believe this paper contributes is to the study of research technology as an input into knowledge production. In particular, we document that the availability of this motion-sensing research technology enables gains in production and movement in ideas space. Further, we demonstrate that the ability to leverage research technology varies across researcher types. While the Kinect shock supports some benefits among within-area researchers, its greatest impact is among researchers who have not previously engaged in motion-sensing research. Acemoglu (2002, 2012) and Bryan and Lemus (2017) worry that economic incentives lead to an under-provision of research diversity that is especially pronounced among for-profit firms and that can be alleviated by increasing research diversity among academics. This paper suggests that automating research technology can lessen this problem by enabling wider exploration of research ideas.

A third contribution we seek to make in this paper involves the adaptation of machine learning tools to develop generalizable measures of type of knowledge production. In our main specifications, we employ HDP, which we train on paper abstracts to parse fields of research into multiple categories. The approach has several advantages over bibliometric measures for capturing the diversity and trajectory of research output. Furthermore, unlike LDA, one of the most-often used topic modeling tools, HDP identifies the optimal number of topics within a body of text. While the technique cannot automatically track the evolution of topics over time, we address this limitation by augmenting it with a clustering technique, TFIDF cosine similarity, to calculate a cosine vector similarity between the yearly topics generated by the HDP algorithm. We believe that this approach adds to similar early efforts of employing machine learning techniques in developing measures of type of knowledge output (Kaplan and Vakili 2015) and, importantly, represents an advance in measuring *changes* in the type of knowledge output.

Lastly, a fourth contribution we hope to make is to expand the discussion of automation in literature in strategic management and organization studies.[22] Although a substantial and insightful literature on the impact of digital technologies exists, much of it appears in journals focused on information systems and economics, and we hope that this work contributes to expanding attention to these topics in core strategy and management research.

One of the advantages of our context is that the peculiar surprise associated with the Kinect technology supports a closer to causal analysis of its impact on the community of researchers in computer science and electrical engineering research. There are numerous analogues outside these areas in which specific research tasks have been automated, including the automation of statistical analysis via tools such as SAS, Stata, and R and the automation associated with combinatorial chemistry techniques. By focusing on a relatively narrow research area and an unanticipated increase in the availability of a key automating technology in that area, we are able to get closer to obtaining causal identification, albeit at the cost of some external validity. Our approach is also limited in disentangling the effects of automation from those of publicity of the technology; the launch of Kinect and subsequent hype were highly public. Moreover, Kinect is an automating technology that reduces the cost of executing the tasks it automates with minimal adjustments or coinvention costs. This differs from other technologies, such as mainframes, servers, and cloud computing (Jin and McElheran 2018), where the automation effects might manifest in more complex ways. In addition, Kinect was offered at a lower monetary cost than previous generations of motion-sensing technology. It is possible that the magnitude of the effects we observe are in part explained by these circumstances.

Overall, our analysis underscores the important role of IT-based research technology for innovation. Understanding this relationship should be of central interest for research-oriented organizations focused on being ahead of competitors in knowledge production. Our study suggests that both access to research technology and the type of knowledge workers employed by these organizations matter. In addition to productivity benefits, automating research technologies may induce benefits of greater research diversity among innovation-focused firms. Whereas absorptive capacity has focused on the complementarity between internal research and development (research and development) efforts and external knowledge, this paper highlights the fact that expertise can be embedded in a research technology and suggests that externally developed research technologies may have effects on productivity and idea mobility within firms. Furthermore, the magnitude of impact of research technology in knowledge production draws attention to the role of market power in technology development and retailing, as well as the complex implications of technology development and pricing strategies. For example, offers of discounted technologies could lead to indirect returns in the form of accelerated rates of innovation. Thus, while our analysis has been conducted at the level of individual researchers, we hope the paper invites continued work on the impact of automating technologies on innovating firms and sectors and lights a path toward such investigations.

Acknowledgments

The authors thank Ajay Agrawal, Kevin Bryan, Avi Goldfarb, Danielle Li, Olav Sorenson, Scott Stern, and Martin Watzinger for substantive discussions of the analysis; three anonymous referees and seminar participants at numerous conferences and seminars for constructive feedback. The authors also thank Aparna Anand, Avik Basu, Robyn Johnson, and Lindsay Raymond for excellent research assistance. The authors are listed in alphabetical order. Both authors contributed equally, and all errors are their own.

Endnotes

[1] Throughout this paper, we use the terms *automating technology*, *IT-based technology*, and *research technology* and similar variants interchangeably to refer to IT-based devices (including both software and hardware) that are used in the process of research and that automate certain research tasks, for example, data-collection tasks and data-analysis tasks.

[2] We consider *diversification* to be the breadth of a researcher's publication portfolio at each point in time t and *trajectory* to be the distance in knowledge space between a researcher's portfolio of projects at times $t-1$ and t.

[3] In one illustrative and compelling example, Bessen (2015) describes the impact of automatic teller machines (ATMs) on bank tellers and bank branches between the 1980s and 2000s. He notes that, on the one hand, ATMs constituted a cheaper, strict substitute for the tasks provided by bank tellers and thus enabled the closing of some locations and a reduction in the number of tellers per branch. On the other hand, however, the diffusion of ATMs induced net gains in bank teller employment by enabling bank branches to be opened in new locations and by enabling tellers to focus on tasks in which ATMs were poor substitutes, such as relationship banking.

[4] This requires the additional assumption that the standards for quality for publication in the domain of automation do not change in response to the automation. We anticipate that shifting standards in the affected the domain would be a second-order impact of research tool automation, though we hope that this issue will be examined more fully by further research.

[5] The Kinect system comprises a color camera, a depth sensor, and a multiarray microphone. These physical features, along with artificial intelligence pattern-recognition software, enabled Kinect to recognize in three dimensions and in real time the movements and facial expressions of multiple people and to acknowledge and distinguish their voice commands (Zhang, 2012). The Kinect system translated this motion-sensing information into actions enabling players to control gameplay.

[6] The data were collected during late 2014. Hence, there is a truncation in the 2014 data coverage, which is evident in Figures 1–8 as a decrement in estimated effects in the year 2014. All results are robust to dropping the 2014 data.

[7] These terms are available on request and are described in greater detail in Teodoridis (2018).

[8] IEEE assigns each publication to 1 of the 51 main categories listed in its taxonomy. We calculate the Euclidean distance based on the percentage of keywords from each category that a researcher collects in his or her publication portfolio between 2007 and 2010. Formally, we calculate

$$DiversificationIndex_i = 1 - \sqrt{\sum_{k=1}^{51} CategoryPercentage_{ik}^2}, \quad (2)$$

where i is the individual researcher, and $CategoryPercentage_{ik}^2$ represents the squared percentage of keywords assigned to researcher i's publications in each of the k main 51 categories of the IEEE taxonomy. Note that by construction, the measure is less than or equal to 1 and never 0, and it increases with higher levels of keywords spread across IEEE categories.

[9] Alternatives to using bibliometric measures to capture movement in the ideas space do exist. For example, Krieger et al. (2018) use the Tanimoto distance, which reflects the similarity of chemical structures, to measure mobility in the ideas space in the pharmaceutical industry.

[10] Machine learning algorithms evolved from the study of pattern recognition in computer science but have increasingly found

applications in a variety of fields, including genetics, medical imaging, computational biology and bioinformatics, image recognition, social network analysis, and economics and public policy (Athey and Imbens, 2015). Currently, a large number of algorithms are customized for various tasks. Although limitations remain, the complexity and accuracy of the algorithms are rapidly evolving.

[11] The HDP is a more advanced version of the more well-known latent Dirichlet allocation (LDA).

[12] Despite substantial advances in computing power, each process is computationally intensive. Each run of HDP using our data requires days of computing time. As a result, we have run the current analysis allocating papers to topics using article abstracts but not full texts. In addition, the large volume of data in our data set prevents us from running the HDP or LDA algorithm on multiple years of data at once to, for example, compare the structure generated by these algorithms with the IEEE taxonomy that covers the entire corpus of publications across all years.

[13] We consulted with experts in computer science, electrical engineering, and electronics to ensure that the topics identified by the HDP algorithm reflect credible categorizations in this line of research.

[14] All our results remain robust to using the LDA algorithm with 40, 60, and 90 topics. We chose the number of topics based on the optimal number of topics identified by the HDP algorithm, cross-checked against a sensitivity algorithm on multiple instances of LDA. We do not include these results because of space limitations but are happy to provide them on request.

[15] Computer scientists are working on a variety of extensions of these algorithms. We chose to use algorithms that are considered robust among computer scientists rather than current experimental ones aimed at advancing the frontier in topic modeling.

[16] All core results are robust to altering these cutoff choices and to including the full set of data available to us. Specifically, the results remain robust (1) to considering the full data set, 2001–2014, (2) to considering other cutoff points for the minimum and maximum number of publications, and (3) to eliminating cutoffs for the minimum and maximum number of publications and using the full set of 2007–2014 authors.

[17] Our data set is a balanced panel where the productivity in non-publishing years is taken into account as years with zero publications.

[18] The publication boost is consistent with that in other work, which examines shocks to the availability of research tools. For example, Furman and Stern (2011) find that making life science research materials available through public resource collections, biological resource centers (BRCs), induces between 50% and 125% increase in research referencing these materials. Similarly, Murray et al. (2016) find that open access to certain types of research mice yields a 22% to 43% boost in research citing the use of such mice.

[19] In addition to observing these differences across estimations, we perform a CEM procedure on our main sample that includes both within-area and outside-area researchers. We add both interaction terms in the same estimations to better compare the magnitude of the effects and observe that indeed the impact on within-area researchers was smaller in magnitude than the impact on outside-area researchers (see Table C1 in Online Appendix C).

[20] As before, we include more formal estimates of the difference in magnitude in Table C3 of Online Appendix C.

[21] Data limitations prevent us from providing comprehensive evidence at the individual researcher level.

[22] We are grateful to an anonymous reviewer for proposing this point.

References

Acemoglu D (2002) Directed technical change. *Rev. Econom. Stud.* 69(4):781–809.

Acemoglu D (2012) Diversity and technological progress. Stern S, Lerner J, eds. *The Rate and Direction of Inventive Activity Revisited* (University of Chicago Press, Chicago), 319–356.

Acemoglu D, Autor D (2011) Skills, tasks and technologies: implications for employment and earnings. Card D, Ashenfelter O, eds. *Handbook of Labor Economics*, vol. 4, (Elsevier, Amsterdam, Netherlands), 1043–1171.

Acemoglu D, Restrepo P (2018) Modeling automation. *Amer. Econom. Rev.* 108:48–53.

Aghion P, Dewatripont M, Stein CJ (2008) Academic freedom, private-sector focus, and the process of innovation. *RAND J. Econom.* 39(3):617–635.

Agrawal A, Goldfarb A, Teodoridis F (2016) Understanding the changing structure of scientific inquiry. *Amer. Econom. J. Appl. Econom.* 8(1):100–128.

Agrawal A, Gans J, Goldfarb A, eds. (2018) *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Chicago).

Argyres NS (1999) The impact of information technology on coordination: evidence from the B-2 'stealth' bomber. *Organ. Sci.* 10(2):162–180.

Athey S, Imbens G (2015) *Lectures on Machine Learning* (National Bureau of Economic Research, Cambridge, MA).

Autor DH, Katz LF, Krueger AB (1998) Computing inequality: have computers changed the labor market? *Quart. J. Econom.* 113(4): 1169–1213.

Azoulay P, Furman JL, Krieger JL, Murray F (2015) Retractions. *Rev. Econom. Stat.* 97(5):1118–1136.

Bessen J (2015) How computer automation affects occupations: technology, jobs, and skills. Working paper, Boston University, Boston.

Bilton N (2011) Microsoft sells 10 million Kinects, breaking record. *New York Times Blogs* (March 9), https://bits.blogs. nytimes.com/2011/03/09/microsoft-sells-10-million-kinects-breaking-record.

Borjas G, Doran K (2012) The collapse of the Soviet Union and the productivity of American mathematicians. *Quart. J. Econom.* 127(3):1143–1203.

Borjas G, Doran K (2015) Cognitive mobility: native responses to supply shocks in the space of ideas. *J. Labor Econom.* 33(S1): S109–S145.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2017) Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science. *Management Sci.* 62(10):2765–2783.

Bloom N, Sadun R, Van Reenen J (2012) Americans do IT better: US multinationals and the productivity miracle. *Amer. Econom. Rev.* 102(1):167–201.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: firm-level evidence. *Quart. J. Econom.* 117(1):339–376.

Bryan K, Lemus J (2017) The direction of innovation. *J. Econom. Theory* 172:247–272.

Brynjolfsson E, Hitt L (2003) Computing productivity: firm-level evidence. *Rev. Econom. Stat.* 85(4):793–808.

Brynjolfsson E, Mitchell T, Rock D (2018) Why can machines learn and what does it mean for occupations and the economy. *Amer. Econom. Rev.* 108:43–47.

Buntine W, Jakulin A (2004) Applying discrete PCA in data analysis. Chickering M, Halpern J, eds. *Proc. 20th Conf. Uncertainty Artificial Intelligence* (AUAI Press, Arlington, VA), 59–66.

Carmody T (2010a) Adafruit offers $1000 bounty for open-source Kinect drivers. *Wired Magazine* (November 4), https://www.wired.com/2010/11/adafruit-offers-1000-bounty-for-open-source-kinect-drivers/.

Carmody T (2010b) How Microsoft learned to stop worrying and love Open Kinect. *Wired Magazine* (November 22), https://www.wired.com/2010/11/open-kinect/.

Choudhury P, Starr E, Agarwal R (2018) Machine learning and human capital: experimental evidence on productivity complementarities. Working paper, Harvard Business School, Boston.

Cockburn I, Henderson R, Stern S (2017) The impact of artificial intelligence on innovation: an exploratory analysis. Agrawal A, Gans J, Goldfarb A, eds. *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, Chicago).

Cohen WM, Klepper S (1992) The tradeoff between firm size and diversity in the pursuit of technological progress. *Small Bus. Econom.* 4(1):1–14.

Cohen WM, Klepper S (1996) A reprise of size and R&D. *Econom. J.* 106(437):925–951.

Cohen WM, Sauermann H, Stephan PE (2018) Academics' motives, opportunity costs, and commercial activities across fields. Working paper, Duke University, Durham, NC.

Deming DJ (2017) The growing importance of social skills in the labor market. *Quart. J. Econom.* 132(4):1593–1640.

Dean, JW, Yoon SJ, Susman GI (1992) Advanced manufacturing technology and organization structure: empowerment or subordination? *Organ. Sci.* 3(2):203–229.

Ding WW, Levin SG, Stephan PE, Winkler AE (2010) The impact of information technology on scientists' productivity, quality and collaboration patterns. *Management Sci.* 56(9):1439–1461.

Dos Santos B, Peffers K (1995) Rewards to investors in innovative information technology applications: first movers and early followers in ATMs. *Organ. Sci.* 6(3):241–259.

Felten E, Raj M, Seamans R (2018) A method to link advances in artificial intelligence to occupational abilities. *AEA Papers & Proc.* 108:54–57.

Forman C, McElheran K (2012) Information technology and the boundary of the firm: evidence from plant-level data. Working paper, Harvard Business School, Boston.

Furman JL, Stern S (2011) Climbing atop the shoulders of giants: the impact of institutions on cumulative knowledge production. *Amer. Econom. Rev.* 101(5):1933–1963.

Furman JL, Murray F, Stern S (2012) Growing stem cells: the impact of federal funding policy on the U.S. scientific frontier. *J. Policy Anal. Management* 31(3):661–705.

Giles J (2010) Inside the race to hack Microsoft's Kinect. *New Sci.* 208(2789):22–23.

Griliches Z (1990) Patent statistics as economic indicators: a survey. *J. Econom. Lit.* 28(4):1661–1707.

Grupp H (1990) The concept of entropy in scientometrics and innovation research. *Scientometrics* 18(3–4):219–239.

Han J, Shao L, Xu D, Shotton J (2013) Enhanced computer vision with Microsoft Kinect sensor: a review. *IEEE Trans. Cybern.* 43(5): 1318–1334.

Hofmann T (1999) Probabilistic latent semantic analysis. Atkinson K, ed. *Proc. 15th Conf. Uncertainty Artificial Intelligence* (AUAI Press, Arlington, VA), 289–296.

Iacus SM, Porro K (2011) Multivariate matching methods that are monotonic imbalance bounding. *J. Amer. Statist. Assoc.* 106(493): 345–361.

Iacus SM, Porro K (2012) Causal inference without balance checking: Coarsened exact matching. *Political Anal.* 20(1):1–24.

Jin W, McElheran K (2018) Economies before scale: survival and performance of young plants in the age of cloud computing. Working paper, Rotman School of Management, University of Toronto, Toronto.

Jones B (2009) The burden of knowledge and the death of the renaissance man: is innovation getting harder? *Rev. Econom. Stud.* 76(1):253–281.

Jones B (2010) As science evolves, how can science policy? Lerner J, Stern S, eds. *NBER Innovation Policy and the Economy* (National Bureau of Economic Research, Cambridge, MA), 103–131.

Kaplan S, Vakili K (2015) The double-edged sword of recombination in breakthrough innovation. *Strategic Management J.* 36(10): 1435–1457.

Karabarbounis L, Neiman B (2014) The global decline of the labor share. *Quart. J. Econom.* 129(1):61–103.

Krieger J, Li D, Papanikolou D (2018) Developing novel drugs. Working paper, Harvard Business School, Boston.

Merton RK (1938) Science and the social order. *Philos. Sci.* 5(3): 321–337.

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

Mokyr J, Vickers C, Ziebarth NL (2015) The history of technological anxiety and the future of economic growth: is this time different? *J. Econom. Perspect.* 29(3):31–50.

Murray F, Aghion P, Dewatripont M, Kolev J, Stern S (2016) Of mice and academics: examining the effect of openness on innovation. *Amer. Econom. J. Econom. Policy* 8(1):212–252.

Murray F, O'Mahony S (2007) Exploring the foundations of cumulative innovation: implications for organization science. *Organ. Sci.* 6(18):1006–1021.

Myers K (2018) The elasticity of science. Working paper, Harvard University, Boston.

Nelson AJ (2016) How to share 'a really good secret': managing sharing/secrecy tensions around scientific knowledge disclosure. *Organ. Sci.* 27(2):265–285.

Nordhaus WD (2007) Two centuries of productivity growth in computing. *J. Econom. Hist.* 67(1):128–159.

Pinsonneault A, Kraemer KL (2002) Exploring the role of information technology in organizational downsizing: a tale of two American cities. *Organ. Sci.* 13(2):191–208.

Rafibakhsh N, Gong J, Siddiqui MK, Gordon C, Felix Lee H (2012) Analysis of Xbox Kinect sensor data for use on construction sites: depth accuracy and sensor interference assessment. Cai H, Kandil A, Hastak M, Dunston P, eds. *Construction Research Congress 2012: Construction Challenges in a Flat World* (ASCE Publishing, Reston, VA), 848–857.

Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics* 82(2):263–287.

Richards-Rissetto H, Remondino F, Agugiaro G, von Schwerin J, Robertsson J, Girardi G (2012) Kinect and 3D GIS in archaeology. Guidi G, Addison AC, eds. *18th International Conf. Virtual Systems and Multimedia*, (IEEE, Piscataway, NJ), 331–337.

Smith A (1776) *An Inquiry into the Nature and Causes of the Wealth of Nations* (W. Strahan and T. Cadell, London).

Teodoridis F (2018) Understanding team knowledge production: the interrelated roles of technology and expertise. *Management Sci.* 64(8):3625–3648.

Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* 101(476):1566–1581.

Terdiman D (2010) Bounty offered for open-source Kinect Driver. CNET News (November 4), https://www.cnet.com/news/bounty-offered-for-open-source-kinect-driver/.

Thompson N, Zyontz S (2017) Who tries (and who succeeds) in staying at the forefront of science: evidence from the DNA-editing technology CRISPR. Working paper, MIT Sloan School of Management, Cambridge, MA.

Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342(6157):468–472.

Volkoff O, Strong D, Elmes M (2007) Technological embeddedness and organizational change. *Organ. Sci.* 18(5):832–848.

Wortham J (2010) Off the shelf and into the garage. *New York Times* (November 22).

Zhang Z (2012) Microsoft Kinect sensor and its effect. *IEEE Multimed.* 19(2):4–10.

Jeff Furman is an associate professor of strategy and innovation at Questrom School of Business, Boston University, and a faculty research fellow at the National Bureau of Economic

Research (NBER). He received his PhD from MIT Sloan School of Management. His recent projects examine the impact of institutions on cumulative innovation, the strategic management of science-based firms, and science and innovation policy.

Florenta Teodoridis is an assistant professor of management and organization at the Marshall School of Business, University of Southern California. Her research areas are the economics of innovation and creativity and the impact of technology on society, with a particular focus on factors influencing the process of knowledge production at the individual researcher level and the relationship between knowledge advancement and markets.