

## Information Technology: Doctoral Theses

### **“Essays on Information Technology, Human Capital, and the Future of Work ”**

**Author:** Sebastian Steffen (2022)

**Committee:** Erik Brynjolfsson (chair), John Horton, Sinan Aral, Prasanna Tambe (The Wharton School, University of Pennsylvania)

**Abstract:**

This dissertation contains three essays concerning the economics of information technology, human capital, and the future of work. In the first essay, 'Occupational Change: Automation and Reskilling Risks', I develop a methodology to study occupational skill demands and estimate the returns to skills, by leveraging novel data from over 200 million online job postings from 2010 until 2020. I find large heterogeneity in skill returns across industries and identify potential (re)skill investment opportunities for workers.

In the second essay, 'Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance', I build on the methodology and data from the previous chapter to measure how feasible it is for firms to shift their workforce to remote work. Using these data, I then causally identify how much remote work practices aided firms' resilience against the Covid-19 pandemic, as measures by sales, net income, stock market returns, and volatility. The findings highlight that firms need to strategically manage the labor composition and digitization of their organizations, and consider that work-from-home practices, besides their many other advantages, are an effective way to hedge against operational risks.

In the final essay, 'Treating the Symptoms or the Cause? Substantive and Symbolic Talent Acquisition in Response to Data Breaches', I use the data from the first chapter to study firms' hiring responses to data breaches. Advancing the theory of substantive and symbolic IT adoption to complementary human capital acquisitions, I find that firms significantly increase their hiring for cybersecurity as well as public relations and legal workers after suffering breach. I also find that public scrutiny can serve as an effective mechanism to shift firms' hiring investments toward substantive, rather than symbolic measures. Given the increase in the volume and severity of cyberattacks, these results provide important and timely insights into firms' responses and incentives to more substantively safeguard their data.

### **“Essays on the Design of Online Marketplaces and Platforms”**

**Author:** David Holtz (2021)

**Committee:** Sinan Aral (chair), Dean Eckles, John Horton

**Abstract:**

In Chapter 1, I estimate the impact of increasing the extent to which content recommendations are personalized. By analyzing the results of a randomized experiment on approximately 900,000 Spotify users across seventeen countries, I find that increasing recommendation personalization increased the number of podcasts that Spotify users streamed, but also decreased the individual-level diversity of Spotify users' podcast consumption and increased the dissimilarity between the podcast consumption patterns of different users across the population. Additional analysis shows that exposure to more personalized recommendations affected not only algorithmically-driven content consumption, but also the content that users consumed organically. The shifts in consumption diversity I observe can affect user retention and lifetime value, and impact the optimal strategy for content producers. These results indicate that personalized recommendations have the potential to create an "engagement-diversity trade-off" when firms optimize solely for consumption.

In Chapter 2, I propose methods for obtaining unbiased estimates of the total average treatment effect (TATE) when conducting experiments in online marketplaces, and test the viability of said methods using a simulation built on top of scraped data from Airbnb. The baseline approach to experimentation -- an individual-level, Bernoulli-randomized experiment analyzed using the difference-in-means treatment effect estimator -- is likely to yield biased TATE estimates when used in online marketplaces, due to, e.g., competition between

sellers in the marketplace. The methods proposed in this chapter, such as graph cluster randomization and exposure modeling, draw on the existing literature on experimentation in networks, and depend on modeling the market as a network, in which an edge exists between two items if they might complement or substitute for one another. I find that blocked graph cluster randomization can reduce the bias of TATE estimates in online marketplaces by as much as 64.5%, however, this reduction in bias comes with a substantial increase in root-mean-square error (RMSE). I also find that fractional neighborhood treatment response (FNTR) exposure models and inverse probability-weighted estimators have the potential to further reduce bias, depending on the choice of FNTR threshold. These results are robust across different treatment interventions, outcomes, levels of network mis-specification, clustering approaches, market structures, levels of demand, and data generating processes.

In Chapter 3, I conduct two large-scale meta-experiments on Airbnb in an attempt to estimate the actual magnitude of bias in TATE estimates from marketplace interference. In both meta-experiments, some Airbnb listings are assigned to experiment conditions at the individual-level, whereas others are assigned to experiment conditions at the level of clusters of listings that are likely to substitute for one another. The two meta-experiments measure the impact of two different pricing-related interventions on Airbnb: a change to Airbnb's fee policy, and a change to the pricing algorithm that Airbnb uses to recommend prices to sellers. Analysis of the fee policy meta-experiment reveals that at least 32.60% of the treatment effect estimate in the Bernoulli-randomized meta-experiment arm is due to interference bias. I also find weak evidence that the magnitude and/or direction of interference bias in online marketplaces depends on the extent to which a market is supply- or demand-constrained. Analysis of the pricing algorithm meta-experiment does not produce a statistically significant estimate of the magnitude of TATE estimate bias due to marketplace interference, but does highlight the difficulty of detecting interference bias when treatment interventions require intention-to-treat analysis.

## **“How Should We Measure the Digital Economy?”**

**Author:** Avinash Collis (2020)

**Committee:** Erik Brynjolfsson (chair), Sinan Aral, Catherine Tucker

### **Abstract:**

Gross domestic product (GDP) measures production and is not meant to measure well-being. While many people nonetheless use GDP as a proxy for well-being, consumer surplus is a better measure of consumer well-being. This is increasingly true in the digital economy where many digital goods have zero price and as a result, the welfare gains from these goods are not reflected in GDP or productivity statistics. Chapter 1 proposes a way of directly measuring consumer's economic well-being using massive online choice experiments. It finds that digital goods generate a large amount of consumer surplus that is currently not captured in GDP. For example, the median Facebook user needed a compensation of around \$48 to give it up for a month. Building up on these results, Chapter 2 extends the GDP framework to include welfare gains from new and free goods and construct a new metric called GDP-B, where B stands for benefits. It finds that including the welfare gains from Facebook would have added between 0.05 and 0.11 percentage points to GDP-B growth per year in the US. Chapter 3 proposes a way of measuring network effects on multi-sided platforms using choice experiments. It also models digital platforms allowing for heterogeneity in demand elasticity and network effects across users of different types. It then calibrates the model using an empirical application to Facebook and simulates six different taxation and regulatory policies. Chapter 4 looks at the impact of social media on subjective well-being and academic performance through a randomized controlled trial of University students. Chapter 5 summarizes the research agenda moving forward and concludes with a framework for measuring different aspects of well-being in the digital economy.

## **“Essays on Information Technology, Intangible Capital, and the Economics of Artificial Intelligence”**

**Author:** Daniel Rock (2019)

**Committee:** Erik Brynjolfsson (chair), Andrew Lo, Chad Syverson, Prasanna Tambe

### **Abstract:**

This dissertation discusses the role of intangible and technological investments in the dynamics of productivity growth, the valuation of firms, and the employment of human capital. The first essay describes the Productivity J-Curve. General purpose technologies (GPTs) such as AI enable and require significant complementary investments, including business process redesign, co-invention of new products and business models, and investments in human capital. These complementary investments are often intangible and poorly measured in the national accounts, even if they create valuable assets for the firm. We develop a model that shows how this leads to an underestimation of output and productivity in the early years of a new GPT, and how later, when the benefits of intangible investments are harvested, productivity will be overestimated. Our model generates a Productivity J-Curve that can explain the productivity slowdowns often accompanying the advent of GPTs, as well as the follow-on increase in productivity later. We use our model to assess how AI-related intangible capital is currently affecting measured total factor productivity (TFP) and output. We also conduct a historical analysis of the roles of intangibles tied to R&D, software, and computer hardware, finding substantial and ongoing effects of software in particular and hardware to a lesser extent.

The second essay investigates the role of engineering talent in predicting the market value of publicly traded firms, then looks in greater detail at AI Talent. Engineers, as implementers of technology, are highly complementary to the intangible knowledge assets that firms accumulate. This paper seeks to address whether technical talent is a source of rents for corporate employers, both in general and in the specific case of the surprising open-source launch of TensorFlow, a deep learning software package, by Google. First, I present a simple model of how employers can use job design as a tool to exercise monopsony power by partially allocating employee time to firm-specific tasks. Then, using over 180 million position records and over 52 million skill records from LinkedIn, I build a panel of firm-level investment in technological human capital (information technology, research, and engineering talent quantities) to measure the market value of technological talent. I find that on average, an additional engineer at a firm is correlated with approximately \$855,000 more market value. Consistent with that finding, AI-intensive companies rapidly gained market value following the launch of TensorFlow. Using a difference-in-differences approach, I show that the launch of TensorFlow is associated with an approximate increase of 4-7% in firm market value for firms employing workers with AI skills compared to firms without AI talent exposure.

The third essay is about which tasks done by workers in the U.S. economy are Suitable for Machine Learning (SML). Advances in machine learning (ML) are poised to transform numerous occupations and industries. This raises the question of which tasks will be most affected by ML. We present a model of labor demand in the presence of new technology and labor constraints following Autor, Levy, and Murnane (2003). We then apply the rubric evaluating task potential for ML in Brynjolfsson and Mitchell (2017) and extended in Brynjolfsson, Mitchell, and Rock (2018) to build measures of “Suitability for Machine Learning” (SML) and apply it to 18,112 tasks in O\*NET. We find that ML has the potential to affect many occupations in the economy, though few (if any) jobs can be completely automated by ML. We discuss the distribution of sensitivity to ML technologies across regions and industries, finding that the effects of ML will follow different patterns than earlier waves of automation.

The fourth essay discusses shifts in the employment of routine labor. A large literature has documented occupational shifts in the US away from routine intensive tasks. Theories of skill-biased technological change differ in whether they predict changes in occupational mix within firms, or merely across different firms or industries. Using LinkedIn resume records, BLS OES data, and Compustat employee counts, we estimate occupational employment for publicly traded US firms from 2000 through 2016. We find that faster employment growth among firms that disproportionately employ non-routine workers is the most important cause of SBTC, followed by within firm occupational mix rebalancing. The entry of new firms also plays a role, although firm exit is slightly routine-worker biased. R&D leads firms to have a larger share of routine workers. These results are most consistent with a theory of routine task demand reduction caused by the diffusion of infra-marginally

implemented new technologies. We also introduce a new measure of business labor dynamism, capturing the frequency with which firms change their occupational mix. Consistent with trends in productivity and other measures of business and labor market dynamism, this measure has decreased steadily since 2000