UNAWARE OR UNACCEPTING: HUMAN BIASES AND THE ARTIFICIAL INTELLIGENCE AVALANCHE

by

Nathan Lane Meikle

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Administration

David Eccles School of Business

The University of Utah

May 2018

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Nathan Lane Meikle**

has been approved by the following supervisory committee members:

Bryan Bonner	, Chair	**3/15/2018** Date Approved
Jack Brittain	, Member	**3/15/2018** Date Approved
Juliana Schroeder	, Member	**3/16/2018** Date Approved
Harris Sondak	, Member	**3/15/2018** Date Approved
Elizabeth Tenney	, Member	**3/15/2018** Date Approved

and by **William Hesterly** , Chair/Dean of

the Department/College/School of **David Eccles School of Business**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Accurately assessing current and future technological capabilities can provide organizations with competitive advantage. However, do accurate assessments become more difficult when technology progresses at a rapid rate? This dissertation examines the rapid growth of artificial intelligence (AI), its impact on organizations, and the human biases that impede accurate assessments of rapid technological progress. Given the magnitude with which AI is impacting organizations, the importance of accurately assessing technological capabilities is hard to overstate. However, I argue that due to the exponential growth bias (i.e., the tendency to underestimate exponential growth), people underestimate both how quickly AI has advanced and is likely to advance. Moreover, due to the motivated reasoning bias (i.e., the desire to search for and interpret information in ways consistent with one's desires), I predict people will reject the aversive implications of rapid technological progress, further impeding accurate assessment. I examine these phenomena through a series of surveys and experiments.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND THEORETICAL BACKGROUND

An accurate assessment of current and future technological capabilities can provide organizations with competitive advantage (Porter, 1985). However, managers who are unaware of current technological capabilities will be hard-pressed to adopt new technology in a timely fashion, let alone benefit from it (Porter & Millar, 1985). Moreover, when adoption of technology requires significant lead time (e.g., in the case of technologies requiring new laws or infrastructure to be in place before adoption), an accurate prediction of technological capabilities can become even more important due to the first mover advantage (Liebermann & Montgomery, 1988) enjoyed by those who exploit technological capabilities.

When technology progresses nonlinearly (e.g., at a continual, exponential, positive rate), an accurate assessment of technological capabilities is challenging for two reasons. First, people consistently underestimate how quickly exponential growth curves accelerate (Stango & Zinman, 2009a; Wagenaar & Sagaria, 1975), a tendency that is called the *exponential growth bias*. The central finding regarding the exponential growth bias is that people persistently and substantially underestimate exponential growth (Stango & Zinman, 2009b).

Second, exponential progress in technology may result in aversive outcomes (e.g., mass unemployment, reduced autonomy, or existential risks; Bostrom, 2014; Ford, 2015; Frey & Osborne, 2017), further impeding an accurate assessment of technological progress. Research in social psychology indicates that conclusions that violate people's desired preferences (e.g., aversive outcomes) will tend to be discounted (Buehler, Griffin, & MacDonald, 1997; Mullen & Skitka, 2006; Paharia, Vohs, & Deshpandé, 2013). This tendency for people to arrive at conclusions they prefer, as opposed to conclusions that are accurate (Dawson, Gilovich, & Regan, 2002; Kunda, 1990; Rousseau & Tijoriwala, 1999) is termed the *motivated reasoning bias* (Kunda, 1990). In numerous domains, individuals' predictions of what will happen coincide with what they would like to see happen (Johnson & Sherman, 1990; Kunda, 1990; Taylor & Brown, 1988). Although people are motivated to make accurate assessments (Petty & Cacioppo, 1986), people are also motivated to arrive at conclusions they prefer (Kunda, 1987, 1990).

This dissertation argues that due to the exponential growth bias (Stango & Zinman, 2009b), people are unaware of, and unprepared for, upcoming changes to organizations stemming from rapid technological progress. Moreover, due to the motivated reasoning bias (Kunda, 1990), people are prone to reject the aversive implications of rapid progress. Thus, I predict people will inaccurately assess current and future technological progress when technological progress is rapid.

To test these predictions, I examine the exponential progress of technology. I explore how technology is changing and will likely change organizations in the future. I argue that people are unaware of, and unprepared for, these changes due to the exponential growth bias. I then discuss some of the potential aversive implications of exponential technological progress (e.g., mass unemployment, computers surpassing human capabilities, computers posing existential risks to humans) and argue that, due to the motivated reasoning bias, people reject aversive implications of rapid technological progress, thus further impeding an accurate assessment of technological capabilities.

Exponential Growth

A simple thought experiment illustrates how quickly technology is advancing, and thus how quickly organizations are both changing and are expected to change in the coming decades. Imagine a person living in the 1700s, who travels primarily by foot and lives in a home heated by fire (Derry & Williams, 1993). Now imagine that person traveling back in time 3,000 years. Presumably, she would not be surprised to be traveling primarily by foot and living in a home heated by fire. Now imagine that same person, who lives in the 1700s. What if she were to travel just *300* years into the future (as opposed to 3,000)? Presumably she would be surprised by the running water, heating and air-conditioning systems, flushing toilets, electrical outlets, radios, telephones, televisions, computers, the Internet, satellites, interstate highways, self-driving cars, jet airplanes, commercial farms, grocery stores, and skyscrapers.

Why the drastic change in technology? The simple explanation is that technology has progressed at an exponential rate (Brynjolfsson & McAfee, 2014; Kurzweil, 2005). For thousands of years, humans correctly expected their future to be much like their past. Exponential growth trends existed, but the progress was so slow as to be imperceptible within one’s lifetime or even across many generations. That is the nature of exponential growth—it starts out relatively slow before accelerating upward rapidly (McKenzie & Liersch, 2011; Stango & Zinman, 2009a).

Exponential growth can lead to outcomes that are difficult to anticipate or appreciate given that humans tend to underestimate exponential growth curves (de Langhe, Puntoni, & Larrick, 2017). To illustrate the limits of intuition when it comes to exponential growth, imagine a disease that spreads by a factor of two, each day. For example, on day one, one person is infected; on day two, two people are infected; on day three, four people are infected. How many days would it take for the disease to spread to 7.5 billion people? The calculations are simple, yet intuitions are often wrong (McKenzie & Liersch, 2011; Stango & Zinman, 2009a). After 27 days, less than 1% of the population would be infected. But just seven days later, the disease would spread 128-fold; every person on earth would be infected.

The tendency to underestimate the growth of exponential curves is called the exponential growth bias (Soll, Keeney, & Larrick, 2013; Stango & Zinman, 2009a). The origins of research on the exponential growth bias can be traced to Wagenaar and Sagaria (1975) and Wagenaar and Timmers (1978, 1979). Their central finding is that people persistently and substantially underestimate exponential growth (Stango & Zinman, 2009b). The bias is robust to different contexts, such as forecasting pollution (Wagenaar & Sagaria, 1975) or forecasting retirement savings growth (McKenzie & Liersch, 2011), and is also robust to different presentations, such as when information is presented in tables and graphs (Wagenaar & Sagaria, 1975) or when information is presented in the form of a story problem (McKenzie & Liersch, 2011).

The exponential growth bias, while often examined in instances of financial returns (Hubbard, Matthews, & Samek, 2016; McKenzie & Liersch, 2011), is especially relevant to technological progress. A clear demonstration of exponential growth in technology is in the field of computing. In 1965, Gordon Moore—who later became the chairman of Intel—noticed that computing capability doubled every two years, while costs simultaneously decreased (Schaller, 1997; Waldrop, 2016). This resulted in an exponential growth in the price-performance of computation (performance increased exponentially as price decreased). For example, in 1968, one dollar could purchase one transistor (a fundamental building block of computing capability), whereas in 2002, one dollar could purchase 10 million transistors; moreover, transistors in 2002 were nearly 1000 times faster than transistors in 1968 (Kurzweil, 2005). In 2017, Intel announced that it could fit more than 100 million transistors in each square millimeter of their chips (Courtland, 2017). In sum, "the critical building blocks of computing: microchip density, processing speed, storage capacity, energy efficiency, download speed, and so on, have been improving exponentially…" (Brynjolfsson & McAfee, 2014, p. 49).

What does this sort of exponential improvement mean for technology products? The Apple iPhone is one such product—it has more computing power than all of NASA's computers combined that helped send humans to the moon (Kaku, 2012; Rosoff, 2014).

Other real-world examples illustrate the effects of exponential growth in technology. In 1990, biochemists set the goal to transcribe the human genome in 15 years; yet, these same scientists had spent the prior year transcribing just one ten-thousandth of the human genome (Venter et al., 2001). At the then-current technological rates, the project would have taken thousands of years to complete; however, due to exponential growth in computing speed (which the scientists accurately predicted), the human genome was decoded just 13 years later (Venter et al., 2001).

One area of technology that is especially benefitting from the exponential progress in computing speed is artificial intelligence (AI; computer systems that are able to perform tasks that normally require human intelligence; Russel & Norvig, 2010). Given that computing capability is the backbone of AI (Brynjolfsson & McAfee, 2014), AIs are expected to improve rapidly, as computers continue to improve. In the following section, I examine recent progress in AI to demonstrate how quickly AIs have progressed in the last several years. I then examine how AIs are changing, and will change, organizations.

Note, when discussing tasks that AIs complete, I use phrases such as *identifying cancer, driving cars, translating languages,* etc. I recognize that AIs do not complete these tasks using the same cognitive processing that humans use. However, for the purposes of this research, I am generally concerned with the outputs of AIs as opposed to the processes by which AIs generate their outputs.

Artificial Intelligence

A 2013 book on judgment and decision making opens with the following sentence: “With little effort, we can accomplish sophisticated tasks, such as recognizing faces or catching a ball, that are far beyond the abilities of even the most powerful computers and sophisticated robots” (Bazerman & Moore, 2013, p.1).

Four years later, the gap between humans and computers has narrowed drastically, as progress on AI “has completely exploded,” (Standage, 2016, p. 2). AIs are identifying cancer (Rose, 2016), driving cars (Levandowski, 2016), trading stocks (Peltz, 2017), translating languages (Marshall, 2017), producing music (Goldhill, 2016), scanning legal contracts (Rosenbaum, 2016), producing news articles (Oremus, 2016), and beating humans in games (Ferrucci, Levas, Bagchi, Gondek, & Mueller, 2013; Silver et al., 2016). For example, in 2011, IBM's Watson beat two of Jeopardy's greatest champions in a multiday event (Ferrucci et al., 2013). Watson was programmed to download all of Wikipedia, and was then able to respond to the similes, jokes, and riddles it encountered on *Jeopardy!* (e.g., when Watson received the following clue in the rhyme category, "a long tiresome speech delivered by a frothy pie topping," Watson correctly responded "A meringue harangue"; Cadwalladr, 2014, para. 25).

In 2015, an AI bested humans in an image recognition challenge (Standage, 2016), and in 2016, Google's DeepMind defeated one of the world's best Go players (Standage, 2016). Google recently announced that their computer can speak nearly as well as a human (van den Oord, Zen, & Dieleman, 2016) and Uber announced that "self-driving" Ubers are available in Pittsburgh (Levandowski, 2016). AIs are also helping companies save energy; Google announced that its AI had achieved a 40% reduction in the amount of energy used for cooling their data centers (Evans & Gao, 2016), a significant feat given the level of sophistication that computer-aided humans had already put in place before the AI produced its solution.

In 2017, the AI Libratus beat four professional poker players in a 20-day tournament. AIs had previously won games in which each player had a full view of a game board (e.g., chess, checkers, go). In poker, however, the AI cannot know what cards the opponent has been dealt nor what decisions the opponent will likely make as a result. In this way, poker more closely mirrors the decisions people face every day (Overly, 2017). To compensate for these uncertainties, a meta-algorithm analyzed the weaknesses in Libratus' strategy at the end of each day, and then algorithmically improved on the weaknesses each night (McLean, 2017). At the end of the tournament, Libratus had won $1,766,250 in hypothetical money, whereas each of the four professionals finished in the negative.

AIs have progressed rapidly and are expected by many to continue to do so (Dobbs et al., 2015). According to some estimates, the "AI avalanche" (i.e., the rapid progress and widespread adoption of AI) will happen faster, and at a broader scale, than the industrial revolution, thus having an even greater impact on humanity (Dobbs et al., 2015; Kurzweil, 2005).

The current wave of AI progress is driven in large part by improved, "deep learning" algorithms (Chou et al., 2017; Jordan & Mitchell, 2015; Standage, 2016). Traditionally, programmers wrote specific instructions for computers to execute, and the computers could only execute those specific instructions (Goodfellow, Bengio, & Courville, 2016; LeCun, Bengio, & Hinton, 2015). For example, if a programmer wanted a computer program to correctly classify an image of a chair, the programmer had to write a list of rules for the program to follow to correctly identify a chair (e.g., does the focal image have four legs, a place to sit, etc.). However, deep learning algorithms rely on vast amounts of data, rather than a comprehensive list of rules, to identify patterns. For example, if a programmer wanted a deep learning algorithm to correctly classify an image of a chair, the programmer would feed the algorithm millions of images of chairs, and let the algorithm determine the patterns that distinguish chairs from other objects (LeCun et al., 2015). Thus, as the deep learning algorithm analyzes more and more data, it improves its capabilities (Goodfellow et al., 2016; Hof, 2013). These improved deep learning algorithms, in conjunction with more powerful computing capabilities, are in large part responsible for the surge in AI capabilities and applications (Chou et al., 2017).

How Is AI Changing Organizations?

Examining the impact that AI is already having on several industries illustrates how consequential AI has become, and foreshadows how consequential AI may soon become. In the following section, I examine several industries—health care, transportation, education, arts and entertainment, retail, finance, and law—that are already changing due to AI advancements.

Health care. Artificial intelligence is changing the way doctors diagnose disease. The University of North Carolina's Lineberger Comprehensive Cancer Center (LCCC) is using IBM's Watson—the same computer system that beat two *Jeopardy!* champions—to diagnose cancer (Rose, 2016). Although scientific research grows at a rate too fast for doctors to consume (some 8,000 academic papers per day), the rate does not surpass a properly resourced AIs' ability (Rose, 2016). The LCCC fed Watson 25 million published medical papers in about one week; Watson also scanned the Web for the latest scientific research. The LCCC then had Watson analyze data from 1,000 cancer patients. In 99% of the cases, Watson identified the same treatments that doctors identified. "The…more exciting part about [the analysis] is in 30% of patients, Watson found something new—so that's 300-plus people where Watson identified [an additional] treatment that a well-meaning, hard-working group of physicians hadn't found" (Rose, 2016, interview on *60 Minutes*).

Other AIs are having success in health care as well. Google researchers are using an eye-scanning AI to detect diabetic retinopathy, which affects almost 1/3 of diabetes patients, and the AI is performing as well as highly trained ophthalmologists (Knight, 2016). AIs are also successfully diagnosing congenital cataracts (Long et al., 2017; Panko, 2017) that when caught early and treated (as they often are in the US) typically do not pose a problem. In the developing world, however, early detection is not common (e.g., in China, roughly 30% of childhood blindness is due to this form of the disease; Panko, 2017). Although human doctors are limited in the number of patients they can see each year, AI systems downloaded as an app on a phone will not have the same limitations.

Bernard Tyson, the CEO of Kaiser Permanente recently stated, “I don’t think any physician today should be practicing without artificial intelligence assisting in their practice. It’s just impossible (otherwise) to pick up on patterns, to pick up on trends, to really monitor care” (Moukheiber, 2014, para. 6). AIs are also being deployed by pharmaceutical companies to speed up drug discoveries (Schubarth, 2017). Moreover, robots that continually improve in dexterity are expected to take on an even larger role in surgeries (Robotics-VO, 2013).

Not only is AI changing the way doctors diagnose disease, AI is also changing the way patients interact with doctors. Veterans in California are participating in virtual psychotherapy sessions, in which they discuss their symptoms with an animated AI avatar on a computer screen (DeVault et al., 2014). According to researchers of the DARPA-sponsored project, participants report overall satisfaction when interacting with the avatar and feel as much rapport with the AI avatar as they feel with face-to-face interviews. Moreover, participants often extend their therapy sessions by double the allotted time (Garber, 2014).

Transportation. In 2015, nearly 40,000 people were killed on American roadways, and another 4.4 million suffered injuries requiring medical attention (National Safety Council, 2016). Worldwide, there are 1.24 million traffic fatalities every year, and 90% of all accidents are due to driver error (Waldrop, 2015). AIs are expected to revolutionize the transportation industry, saving millions of lives in the process. Self-driving cars are already on the roadways in Pittsburgh (Levandowski, 2016) and Washington D.C. (Muoio, 2016), and Google's self-driving car has driven more than 2 million miles. Although it had been in 14 collisions as of 2016, 13 of those accidents were caused by the person driving the other car (Bhuiyan, 2016). Just how soon AIs can drive and fly better than humans, and how quickly humans are willing to adopt self-driving cars and planes, has life and death consequences. "A pilotless airliner is going to come; it's just a matter of when," said Boeing executive James Albaugh in 2011 (Dormehl, 2017, p. 132).

Education. An online AI course at the Georgia Institute of Technology surprised students at the end of the semester—one of the TAs that students had been interacting with all semester (Jill Watson) was actually an AI (Korn, 2016). As AIs improve their ability to understand natural language and answer questions, a more individually tailored education may become the norm. Rather than students receiving the same lecture at the same time and pace, AI-instructors would be programmed to customize lectures, focusing on topics where each individual student is weak. Students would be able to learn at their own pace for presumably a fraction of the cost, given that AIs would not be bound by time (e.g., lectures would be available whenever the students are ready) and space (e.g., students would be able to watch lectures on their computers, rather than in classrooms).

Arts and entertainment. AIs are now capable of producing art and music. For example Sony recently released an AI-produced, Beatles-style pop song (Sony, 2016). The AI was fed a database of 13,000 lead sheets (i.e., scores that record the melodies and harmonies of tracks) from several different styles and composers, and then the researchers programmed the AI to make a song in the style of the Beatles. The song has been played on YouTube 1.5 million times. Google also recently announced a project (Magenta) aimed at using “machine learning to create compelling art and music” (Eck, 2016).

Organizations are now using AIs to produce music as well. Rather than pay musicians to write music for a commercial, companies are now having AIs produce the music at a fraction of the time and cost (Marshall, 2017). For example Jukedeck, a company that sells music produced by AI, recently announced that their AI had produced nearly 1 million songs for customers (Gregory, 2016). Music, however, is not the only entertainment field affected by AI. An AI recently produced a science fiction screenplay after being fed scripts of dozens of science fiction movies (e.g., *Ghostbusters*, *Interstellar*, *The Fifth Element*; Hal, 2016).

Entertainment companies are also incorporating AI into their products. For example, video games depend on AIs to generate the behaviors of the nonplaying characters in the games (Microsoft, 2009). Virtual reality systems (VR) also take advantage of AI. When using VR, users view immersive, three-dimensional environments. The VR system tracks head and eye movements, and changes the images displayed correspondingly, making the users feel part of the world they are viewing on the display.

Retail. The retail industry has undergone significant changes over the last several decades due to information technologies, and AI is expected to continue altering the industry (Whitler, 2016). The last several years have seen a massive shift to online shopping, allowing companies such as Amazon to collect vast amounts of data on each of its customers (Whitler, 2016). These data allows for customized marketing to a degree that was never before possible (Whitler, 2016). For example, AIs predict (via data analysis) buying behavior, and thus make customized recommendations to consumers. AIs, by analyzing weather, purchase rates, and consumer behavior can also help companies manage inventory levels. AIs can also analyze data to help optimize store layouts, enabling fully automated checkout (Chou et al., 2017).

Finance. Whereas banks in the past relied on personal relationships when making lending decisions, banks of the future may rely exclusively on algorithmic decision making. Indeed, some companies already advertise digital mortgage advisors (i.e., algorithms that guide potential buyers through the loan origination process; Davidson, 2106). Stock market trading is also giving rise to algorithmic decision making. More than 50% of all stock market trading is automated, according to some estimates (Phillips, 2013). Hedge funds are also using AIs to make decisions: the company Aidyia has a hedge fund that makes all trades using AI (Metz, 2016). “If we all die” says the company’s chief scientist, “it would keep trading” (Metz, 2016, para. 1). Other traders use AI to instantly capture President Trump’s tweets, for example, and then immediately buy or sell stocks affected by the tweets (Peltz, 2017).

Law. Given AIs’ ability to quickly process vast quantities of documents, law firms of the future will likely be quite different from those of today. AIs can already scan and analyze contract documents with the same or greater accuracy than lawyers in 90% less time (Rosenbaum, 2016). Whereas law firms can generate vast revenues by assigning associates to read through thousands of documents during discovery, AIs can now scan and analyze documents in a fraction of the time, at a fraction of the cost (Sobowale, 2016). Not only are AIs analyzing documents, but also they can be programmed to predict the outcomes of cases (Randazzo, 2016). For example, an AI accurately predicted the outcome of U.S. Supreme Court cases 70% of the time (Katz, Bommarito II, & Blackman, 2017), whereas human experts have been shown to accurately predict cases only 66% of the time (Ruger, Kim, Martin, & Quinn, 2004). AIs are also being programmed to represent clients. One successfully appealed 160,000 parking tickets in 21 months—for free (Gibbs, 2016).

How Will AI Change Organizations?

In September 2016, Stanford released its first report of the "One Hundred Year Study on Artificial Intelligence" (Stone et al., 2016). In the report, the 17-member study panel (made up of AI experts in academia, corporate laboratories, and industry) identified several domains where they expect AI to influence society in the next 15 years: transportation, service robots, healthcare, education, public safety and security, employment and workplace, and entertainment. Although AI will indeed influence each of the aforementioned domains, this dissertation argues that people, due to the exponential growth bias, are unprepared for just how quickly and radically AI will alter the workplace.

Organizational restructuring. One hundred years ago, electricity transformed organizations (Jones, 1991). First of all, it was complicated and was not yet standardized—individual organizations had to decide whether to adopt AC or DC power, which voltages to use, how much to pay, and whether to focus on building lights or replace gas turbines with electric motors, etc. (Ng, 2016). Many organizations responded to these complexities by hiring vice presidents of electricity (Ng, 2016; Rayess, 2015). Twenty years ago, as companies struggled with how to incorporate the Internet into their businesses, companies responded by hiring chief information officers (Synnott & Gruber, 1981). In the same ways that electricity and the Internet changed organizational structures, AI is poised to do the same. Chief AI officers (CAIOs) may soon be as commonplace as chief financial officers, specifically in companies that generate large amounts of data (Ng, 2016). Moreover, we may soon see large companies that are entirely dedicated to AI. For example, Andrew Ng, the former Stanford computer science associate professor, who later founded Coursera and Google Brain, led a 1,300-person AI group at Baidu, the Chinese-American Internet company (Lewis-Kraus, 2016).

Not only will companies be dedicated to managing AIs, but also we may soon see companies produced by AIs themselves. The saying goes that the first ultra-intelligent machine is the last invention that humans will ever make (Good, 1965). If AIs become more capable than humans, it is the AIs that will do the inventing. For example, a recent paper describes how researchers at Microsoft and Cambridge built an algorithm *that writes algorithms* capable of solving simple math problems (Balog, Gaunt, Brockschmidt, Nowozin, & Tarlow, 2017).

As the age of "big data" advances, AIs would be expected to become more and more valuable. Indeed, a key enabler of the technological progress witnessed today is the vast amount of data available, data which AIs analyze and interpret at a speed and efficiency impossible for humans (Dobbs et al., 2015). Whether companies are human based, AI based, or even Internet based, companies that previously did not produce or acquire large amounts of data will likely be doing just that, as more and more everyday objects become networked. The connection of these everyday objects to the Internet is termed the Internet of Things (Xia, Yang, Wang, & Vinel, 2010).

How Will the Internet of Things (IoT) Change Organizations?

The connection of physical things to the Internet allows for physical control from a distance (Kopetz, 2011). *Smart objects* (embedded systems connected to the Internet) are the building blocks of the IoT. When the Internet was first developed, it consisted of only a few nodes. Fifty years later, it has become a pervasive network servicing over 1 billion users (Kopetz, 2011). The rapid progress of technology, including the miniaturization and cost reduction of electronic devices, is now making possible the expansion of the Internet into objects. Each object that is connected to the Internet bridges the gap between the physical world and the information world. For example, consider the 1-35, eight lane, Mississippi River bridge that collapsed in 2007, killing 13 people and injuring 145. Had the bridge been built with *internet-connected cement*, the sensors could have monitored stress, cracks, and warpages, communicating this information to the relevant authorities (Burrus, 2017). Other examples of the benefits of IoT abound: Internet-connected bridges that detect ice could notify smart cars that are approaching the bridge; Internet-connected refrigerators that track the availability and expiry date of food items could autonomously place orders at the grocery store when food supplies are below a predefined limit. However, the real impact of IoT is not felt so much in the functional capability of any individual object, but rather in the collective capability of billions or even trillions of Internet-connected objects (Kopetz, 2011). For example, some analysts predict that by the year 2020, there will already be 50 billion Internet-connected devices (Dormehl, 2017). Whereas the IoT allows for the gathering of vast amounts of data that was never before possible, AI allows for the efficient analysis of that data, due to recent advances such as deep learning, which requires very little engineering by hand, allowing AIs to easily take advantage of increases in available data (LeCun et al., 2015).

Why should organizations care? As the use of AI becomes more prevalent, managers, shareholders, board members, and other stakeholders will decide whether to accept or reject the use of it. Organizations that accurately assess current and future technological capabilities can enjoy competitive advantage. Organizations that adopt AI too soon will bear substantial costs as the technology is refined, whereas organizations that adopt AI too late will miss out on enhanced organizational capability. The case of self-driving cars illustrates the importance of accurately assessing technological capabilities. The moment that self-driving AIs become better drivers than humans is the moment that self-driving AIs should be deployed (and adopted) on a wide scale (if the primary concern is reducing traffic fatalities). If humans defer to self-driving AIs too soon, however, people will needlessly die due to AI error; but if humans defer to self-driving AIs too late, people will needlessly die due to human error. AI adoption decisions will also have far-reaching consequences in industries such as military defense, agriculture e-commerce, education, energy, finance, government, health care, manufacturing, and real estate.

The capability of AI, as it progresses rapidly, has the potential to outpace people's expectations for it—especially if growth rates accelerate. Thus, people would be prone to adopt the use of AI later than is optimal when adoption requires lead-time. For example, a self-driving car that is 34% as capable as a human driver is still far away from equaling human level-ability when assuming linear progress. However, assuming exponential progress, 34% is only one tripling away from equaling human-level performance. Thus, policy makers, manufacturers, drivers, and other stakeholders, thinking about progress linearly, would overestimate the time they have to prepare for the advent self-driving cars. An ability to forecast AI progress (i.e., exponential growth) would thus be required of those decision makers aiming to adopt AI at the optimal time. Moreover, given the accelerating rates of product adoption (Dobbs et al., 2015), and thus the enhanced competitive advantage enjoyed by first-movers, the importance of accurately assessing AI capabilities is hard to overstate.

Proposition 1: People, on average, underestimate the current and future capabilities of AI.

Summary

Rapid technological progress can be difficult to appreciate or anticipate, due to the exponential growth bias. Rapid improvements in AI, a relatively new field (the first AI conference was held in 1956; Dormehl, 2017), have changed and will continue to change organizations. Tasks that were once thought to be exclusively in the domain of human ability (i.e., diagnosing disease, driving cars, matching faces, producing art) are now successfully and routinely completed by AIs. The recent surge in AI progress is due in large part to deep learning algorithms, enhanced computing capability, and vast amounts of data. Numerous industries are already adopting the use of AIs, portending even greater change as AIs continue to improve and become more commonplace. Rapid improvements in AI may lead to organizational restructuring and new organizations altogether. Organizations will increasingly rely on machine decision making rather than human decision making, as vast amounts of data become available due, in part, to the IoT. Organizations that most accurately assess current and future capabilities of technology will enjoy competitive advantage.

Motivated Reasoning

In numerous domains, individuals' predictions of what will happen tend to reflect what they would like to see happen (Buehler et al., 1997; Johnson & Sherman, 1990; Kunda, 1990; Taylor & Brown, 1988). Although people are motivated to make accurate assessments (Petty & Cacioppo, 1986), people are also motivated to arrive at conclusions they prefer (Kunda, 1987, 1990). This desire to search for and interpret information in ways that are consistent with one's beliefs and desires is called motivated reasoning (Kunda, 1990).

A large body of research documents that people self-servingly interpret ambiguous information and reach conclusions consistent with their desires (Babcock & Loewenstein, 1997; Dawson et al., 2002; Gilovich, 1983; Simmons, Nelson, & Simonsohn, 2011; Zuckerman, 1979). Some of the first studies on motivated reasoning demonstrated that people generate theories that view their own attributes as more predictive of desirable outcomes. Consequently, people optimistically believe that good things will happen to them and bad things will not happen to them (Kunda, 1987). For example, people think they are more likely than their peers to experience positive financial and health-related outcomes (Perloff, 1983; Weinstein, 1980, 1983, 1984). People also tend to believe that they are less likely than their peers to be victims of crime (Perloff, 1983). More recently, researchers have found that people are prone to believe in a favorable future in terms of politics, scientific beliefs, and entertainment and product preferences (Rogers, Moore, & Norton, 2017). Regarding people's predictions of AI, a recent Gallup poll found that most Americans think AI will improve lives and eliminate jobs—just not eliminate their jobs (Carlson, 2018).

The prospect of living in a world where machines replace humans in the workforce on a massive scale is likely not appealing to most people (Ryan & Deci, 2000). Although people may willingly accept that AI is progressing rapidly and will continue to do so, research on motivated reasoning (Kunda, 1990) suggests that people will discount the aversive implications of rapid technological progress (e.g., mass unemployment, existential risks, reduced autonomy, etc.). Given that one's motives can affect the beliefs about information to which people are exposed (Buehler, Messervey, & Griffin, 2005; Dawson et al., 2002; Gilovich, 1983; Kappes, Faber, Kahane, Savulescu, & Crockett, 2018; Kunda, 1990; Mullen & Skitka, 2006; Paharia et al., 2013; Rousseau & Tijoriwala, 1999; Simmons et al., 2011), I expect people to resist the idea of human inferiority relative to machines, when inferiority is associated with what people deem to be an aversive implication (e.g., reduced autonomy). The following section explores aversive implications of rapid technological progress, such as unemployment and existential risk.

Aversive Implications of AI Progress

As AIs continue to progress, AI's may soon cause widespread unemployment. When one considers that the basic skills required of family physicians are skills that AIs currently possess, it seems that the age of AI doctors is already upon us. For example, AIs can respond to natural language (e.g., listen to symptoms), process images (e.g., incorporate the information contained in MRIs, CT scans, x-rays), reduce and sort data (e.g., identify pathologies) and prescribe treatments (e.g. Watson prescribing cancer treatments). Moreover, although some people may prefer interacting with human doctors, others may prefer interacting with AIs. Given that AIs can already effectively carry out crucial tasks required of doctors, while potentially reducing embarrassment and insecurity of patients, the medical industry may soon change drastically.

Medicine is not the only industry that will presumably be affected by AIs. As AIs become better drivers (and potentially surpass the ability of humans), the 3.5 million truck, taxi, and bus drivers may be looking for new jobs (Kitroeff, 2016). Given the number of human-caused traffic fatalities each year, it is not hard to imagine a future in which humans are not permitted to drive. Support staff jobs will also likely be affected by AI. As Apple's Siri, Google's Assistant, and Amazon's Alexa become even more proficient personal assistants, millions of support staff workers and customer service representatives may be looking for new jobs as well ("Occupational employment statistics," 2015).

Education seems especially ripe for major change. As AI-instructors are able to customize lectures in ways, and on a scale, that human instructors are incapable of, millions of teachers and professors may be looking for new jobs. If AIs like Watson gain their expertise by downloading information from the Internet, perhaps humans are not far away from being able to rapidly download information directly into their brains (further reducing the need for teachers). Elon Musk's recently-announced company Neuralink apparently aims to help humans do just that—Musk believes that an implantable brain-computer interface is just five years away ("Elon Musk enters the world of brain-computer interfaces," 2017).

Research faculty at universities are also not immune to AI competition. Given that the key strength of AIs is pattern recognition and data analysis, it is possible that AIs will soon be helping humans develop and test theories. Producing a results section of a research paper presumably would not be a problem for AIs, given that AIs are already producing news articles on Minor League Baseball games for the associated press, using only the box score (i.e., data from the games) as input (Oremus, 2016; Press, 2016). In the same way that researchers currently test human-based theories on humans, perhaps researchers will be interested in conducting AI-based theories on AIs.

AI researchers are very interested in how humans treat AIs. Research demonstrates that under some circumstances people make meaningful distinctions when interacting with computers, compared to humans (Blascovich et al., 2002; Gray, Gray, & Wegner, 2007; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; van't Wout, Kahn, Sanfey, & Aleman, 2006; Waytz, Gray, Epley, & Wegner, 2010). For example, in social decision making people do not feel guilty about exploiting machines, but do tend to feel guilty when exploiting humans (Melo, Marsella, & Gratch, 2016). Evidence also suggests that people routinely prefer humans to machines when engaging in social decision making tasks (e.g., prisoner dilemma tasks) (Gallagher, Jack, Roepstorff, & Frith, 2002;

Kircher et al., 2009; Krach et al., 2008; McCabe, Houser, Ryan, Smith, & Trouard, 2001; Rilling et al., 2002; Sanfey et al., 2003; van't Wout et al., 2006). "[R]esearch shows that people can reach different decisions and show different patterns of brain activation with machines in the exact same decision-making tasks, for the exact same financial incentives, when compared to humans" (Melo et al., 2016, p. 8:2). The question arises as to whether and how humans will continue to differentiate between computers and humans as AIs become more commonplace.

When one considers the current capabilities of AIs, and then considers any rate of progress, it becomes progressively harder to think of industries or organizations that could not be drastically altered by AI. According to the market research firm Tractica, the annual global revenue for AI products and services will increase by a factor of 60 over the next nine years, from $600 million to $36 billion (Feldman, 2016).

Although many of the examples above may seem far-fetched, or at the very least far away, a recent study argues that 47% of total U.S. employment is at "high risk" of being automated in the next decade or two (Frey & Osborne, 2017). The authors argue that employees most likely to be affected are those in transportation, logistics, office and administrative support, and service occupations. If indeed AIs replace such a large percentage of workers, and those workers are unable to retrain fast enough to compete in the labor market, there could be widespread unemployment, termed technological unemployment (Ford, 2015). Some researchers are thus calling on politicians to examine universal basic income, a guaranteed minimum income that citizens would receive (Ford, 2015). Finland and Canada recently announced experimental tests of this type (Graham, 2017). If indeed AIs replace human employees on a wide scale, humanity's role in the work place may fundamentally change.

Existential concerns. AI has the potential to solve problems that humans are currently unable to solve, but AI also has the potential to cause immense harm. Stephen Hawking claimed that AI may be the biggest event in the history of our civilization, but it could also be the last (Roberts, 2016). Others share Hawking's concern. Elon Musk claimed, "We need to be super careful with AI. Potentially more dangerous than nukes" (Khatchadourian, 2015, para. 5). Bill Gates echoes Musk's concern, "when people say [AI] is not a problem, then I really start to get to a point of disagreement. How can they not see what a huge challenge this is?" (Khatchadourian, 2015, para. 5).

Simple thought experiments illustrate some of the potential dangers posed by AIs. For example, if an AI had the goal to produce as many widgets as possible, the AI might view humans as an impediment to its goal, and thus eliminate humans. Or perhaps engineers try to constrain AIs with goals that appear to be perfectly safe, for example giving AIs the goal to make humans smile. Without additional constraints, the AI might decide that the most efficient way to make humans smile is to paralyze human facial muscles in the shape of a smile—or stick electrodes in the pleasure centers of the brain, and eliminate all other body parts that are not useful for experiencing pleasure (Bostrom & Müller, 2014). Thus, robots could become fully-automated bureaucrats executing the letter of the law rather than respecting the spirit of the law.

Experts are rightly concerned about the unanticipated and unintended consequences of developing AIs, especially if the AIs eventually surpass human intelligence. In the same way that humans kill insects, often without a second thought (not always due to malicious intent, but simply in defense or a feeling of intellectual superiority), AIs might not be as benevolent to humans as we might hope (Bostrom & Müller, 2014). In response to the existential concerns surrounding AI, the principal U.S. technology companies have recently joined a partnership: Apple, Amazon, Facebook, Google/DeepMind, IBM, and Microsoft have all joined the Partnership on AI, an organization designed "to support research and recommend best practices in areas including ethics, fairness, and inclusivity; transparency and interoperability; privacy; collaboration between people and AI systems; and of the trustworthiness, reliability, and robustness of the technology" (Murphy, 2017, para. 2).

Academia is also taking notice of the potential risks and benefits of AI. The following universities recently announced centers dedicated to the study of AI: Carnegie Mellon (Markoff, 2016), University of Southern California, University of California, Berkeley (Khan, 2016), Cambridge, Oxford, and Imperial College, London (Roberts, 2016). The stated goal for these centers is to ensure that AIs benefit humans. History may later judge these attempts as unevolved, futile, misdirected, or even tyrannical. University departments are also taking notice, although they are struggling to compete with industry in respect to retaining their experts. Uber recently hired 40 people from Carnegie Mellon's National Robotics Engineering Center (Ramsey & Macmillan, 2015). Alphabet Inc. recently hired the director of Stanford University's artificial intelligence lab, and AI researchers from Toronto, NYU, Stanford and Carnegie Mellon recently joined the companies Facebook, Baidu Inc. and Amazon (Hernandez & King, 2016). Although the number of Ph.D.s in the field of computer science has grown, the proportion staying in academia has hit a historic low (Hernandez & King, 2016). Thus, as AI becomes increasingly dominant in industry, relatively fewer and fewer academic computer scientists are available to study the phenomenon.

Proposition 2: People, on average, accept arguments about the rate of technological progress and simultaneously reject the aversive implications of such rapid progress.

Summary

People are self-serving in their interpretation of information and are especially good at arriving at conclusions consistent with their desires (and bad at arriving at undesired conclusions; Kunda, 1990). Moreover, rapid improvements in AI may lead to mass unemployment and even existential threats. Thus, I predict people will reject the aversive implications of rapid progress in AI, even when they are willing to accept that the rate of progress is especially rapid.

CHAPTER 2

RESEARCH METHODOLOGY

AI is advancing rapidly and changing organizations. An accurate assessment of current and future technological capabilities can provide organizations with competitive advantage, but human biases (i.e., exponential growth bias and motivated reasoning bias) can impede people from making accurate assessments. I expect participants to underestimate both how quickly AI has advanced and is likely to advance going forward. Moreover, I expect these errors in judgment to become more pronounced as growth rates accelerate. Lastly, I expect participants to be more likely to reject the implications of rapid progress that they find aversive and to be more likely to accept the implications they find to be desirable. Thus, the goals of the studies in this dissertation are the following:

1. Test the prediction that participants underestimate current AI capabilities.
2. Test the prediction that participants underestimate how quickly AI is likely to progress.
3. Test the degree to which participants underestimate exponential growth.
4. Test the prediction that even when participants accept arguments about technological progress, they often reject the aversive implications of such arguments.
5. Test the prediction that participants reject the aversive, and accept the desirable,

implications of exponential technological growth.

Study 1: Do Participants Underestimate Current AI Capabilities?

Study 1 tests whether participants tend to underestimate current AI capabilities. Given that people drastically underestimate exponential growth rates, I expect participants to underestimate how quickly AI has progressed.

Hypothesis 1: Participants underestimate current AI capabilities.

Sample

Two-hundred thirteen participants[1] in the United States completed the study via Amazon Mechanical Turk in exchange for $0.30 (64.3% female; mean age = 37.7).

Procedures

Participants read the following:

> Assume for the purpose of this question that human scientific activity continues without major negative disruption.[2] In what year would you expect computers/robots to be able to complete the following tasks (or to have already completed the following tasks)? For each of your answers, you will also provide a confidence interval: a numerical range (a lower and upper bound) that you are 90% sure contains the correct answer. If

---

[1]Based on experience with past studies and effect sizes, I generally aimed for a sample of 200 per condition. The exceptions were Study 3 (in which I expected a large effect, and thus aimed for a sample of 75 in each condition) and Study 4 (in which I expected a small effect, and thus aimed for a sample of 400 in each condition).

[2] Prior to running Studies 1 and 2, I conducted a pilot study to examine participants' interpretation of the sentence, "Assume for the purpose of this question that human scientific activity continues without major negative disruption." Of the 32 total respondents, two respondents demonstrated an inadequate grasp of the sentence. This was unsurprising, given the number of participants who routinely fail manipulation-check questions in my MTurk studies.

> you have “no idea” what the answer is, then give a very wide range; and if you happen to be quite certain then give a narrow range. Lastly, if you think computers/robots will never be able to accomplish the task described, write "never" in the final column (and leave the first three columns blank).

Participants then wrote down their answers for each of 12 tasks (presented in random order). In choosing the tasks, I first compiled a mix of intellective tasks (i.e., tasks that have objectively determined best answers; Laughlin, 1980) that computers/robots had clearly completed (i.e., beat a human in an image matching contest, correctly identify types of cancer in cancer patients, beat a professional poker player in a game of poker, and scan legal documents for errors). Next I compiled a mix of judgmental tasks (i.e., tasks that have an ambiguous performance standard; Laughlin, 1980) that computers/robots may or may not have completed, depending on one’s judgment of the AI’s product (i.e., produce music that is indistinguishable from human-created music, as judged by 100 random people, translate languages at a level indistinguishable from human level, as judged by 100 random people, produce computer programs that are comparable to human-created computer programs, as judged by 100 random people, and write news articles that are indistinguishable from human-created news articles, as judged by 100 random people). Finally, I compiled a mix of tasks that computers/robots had clearly not completed (i.e., host a game show, organize a fundraiser, write a blockbuster screenplay, and found a religion). These different categories of tasks allow for an assessment of whether participants tend to overestimate, underestimate or accurately estimate the current capabilities of AI. Lastly, it might seem unusual to use verbs such as *beat*, *identify,* and *scan,* when describing the actions taken by computers. I recognize that computers do not perform tasks with the same cognition that humans do; or at least the cognitive processes of humans and computers are different (if computers have any cognitive processes at all), given the different biological/mechanical make up of humans and computers. However, for the purposes of this research, I am generally concerned with the outputs of computers/robots, as opposed to the processes by which computers generate their outputs.

Results

To assess whether participants, on average, tend to underestimate AI capabilities, I examined the median prediction of all participants to each of the twelve questions (data were collected on September 19, 2017). I report the results segregated by the three categories of questions (i.e., tasks that computers/robots have clearly completed, may or may not have completed, and have not completed). For the first category of responses (Table 1), I also report the year that computers/robots completed the corresponding tasks. Finally, I report the percentage of participants who predicted that computers will never complete the tasks described.

For the first set of tasks (those that computers/robots have already completed), the median response of participants ranged from year 2020-2030 (Table 1). Thus, more than half of participants predicted that computers/robots would complete each of the tasks at some future date. Of note, 5% - 15% of participants predicted that computers/robots will never be able to complete the four tasks that have already been completed by computers/robots.

For the second set of tasks (those that computers/robots may or may not have completed), the median response of participants ranged from year 2025-2030 (Table 2).

Table 1. Median Prediction for Tasks That Computers/Robots Have Completed

Task	Median Prediction of Participants	Year Actually Completed	% of Participants Answering "Never"
Beat a human in an image matching contest	2020	2015	10%
Correctly identify types of cancer in cancer patients	2030	2016	15%
Beat a professional poker player in a game of poker	2020	2017	7%
Scan legal documents for errors	2020	2017	5%

Table 2. Median Prediction for Tasks That Computers/Robots May Have Completed

Task	Median Prediction of Participants	% of Participants Answering "Never"
Produce music that is indistinguishable from human-created music (as judged by 100 random people)	2030	30%
Translate languages at a level indistinguishable from human level (as judged by 100 random people)	2025	15%
Produce computer programs that are comparable to human created programs (as judged by 100 random people)	2030	15%
Write news articles that are indistinguishable from human-created news articles (as judged by 100 random people)	2030	29%

Of note, 15% - 30% of participants predicted that computers/robots will never be able to complete the four tasks that already may have been completed by computers/robots.

For the third set of tasks (those that computers have not completed), the median response of participants ranged from 2030 – “Never” (Table 3). The percentage of participants answering “Never” ranged from 26% - 73%.

The results from Study 1, specifically the first and second segments of tasks, provide support for the hypothesis that participants underestimate current AI capabilities.

Study 2: Do Participants Underestimate How Quickly AIs will Progress?

An ability to foresee future events is required to assess whether people today underestimate future progress. However, a perfect knowledge of the future is presumably unknowable. Thus, Study 2 compares laypeople’s expectations of technological progress with the expectations of AI experts (for examples of studies comparing experts' opinions to laypeople's opinions, see Dryer & Horowitz, 1997; Littlepage, Robison, & Reddington, 1997; Littlepage, Schmidt, Whisler, & Frost, 1995). Expert opinion refers to estimates of people who have exceptional knowledge about some topic (Baum, Goertzel, & Goertzel,

Table 3. Median Prediction for Tasks That Computers/Robots Have Not Completed

Task	Median Prediction of Participants	% of Participants Answering “Never”
Host a game show	2030	26%
Organize a fundraiser	2050	41%
Write a blockbuster screenplay	Never	52%
Found a religion	Never	73%

2011). Study 2 replicates two prior studies that examined experts' opinions regarding AI advancement. However, rather than examine experts' opinions, Study 2 examines laypeople's opinions, and then compares laypeople's responses with experts' responses. *Hypothesis 2: Participants, compared to AI experts, underestimate AI progress.*

Sample

Three-hundred eighty-eight participants in the United States completed the study via Amazon Mechanical Turk in exchange for $0.25 (59% female; mean age = 36.0).

Procedures

Participants completed one of the replications described below:

HLMI replication. Bostrom and Müller (2014) asked 550 AI experts (170 responded) to estimate the time-frame in which high-level machine intelligence (HLMI; i.e., an intelligence that can carry out most human professions as well as a typical human) would be developed. The experts were identified by one of four ways: attendees of the Philosophy and Theory of AI conference, attendees of the Artificial General Intelligence Conference, Members of the Greek Association for Artificial Intelligence, and the Top 100 'Top authors in artificial intelligence' by citation in all years according to Microsoft Academic Search in May 2013. Participants completed the following task:

> Definition: "High–level machine intelligence" is defined as an intelligence that can carry out most human professions at least as well as a typical human. For the purposes of this question, assume that human scientific activity continues without major negative disruption. By what year would you see a (10% / 50% / 90%) probability for such HLMI to exist?"

The median estimate of the AI experts was for a one-in-two chance that high-level machine intelligence will be developed around 2040-2050, rising to a nine-in-ten chance by 2075 (Bostrom & Müller, 2014). I predict that laypeople will overestimate, compared to AI experts, how long it will take for HLMI to exist.

Turing Test replication. Baum and colleagues (2011) asked 21 experts to estimate when a computer will pass the Turing Test (described below). All participants were attendees at the Artificial General Intelligence conference in 2009. Although several versions of the Turing Test exist, participants in Study 2 made estimates based on the one hour test (as opposed to the five minute test) to be consistent with Baum and colleagues (2011). Participants read and answered the following questions (adopted from the Turing Test described in Russel & Norvig, 2010):

> Alan Turing designed a test to determine whether computers could be called intelligent. The test is for a computer program to have a conversation (via online typed messages) with an interrogator for one hour. A computer passes the test if a human interrogator, after posing written questions to the computer, cannot tell whether the written responses come from a person or not. To pass the test, a computer would need to be able to do the following:
> 1. Understand language
> 2. Store what it knows or experiences
> 3. Use the stored information to answer questions and draw new conclusions
> 4. Adapt to new circumstances and detect and extrapolate patterns
>
> Assume for the purpose of this question that human scientific activity continues without major negative disruption. By what year would you see a (10% / 50% / 90%) probability that computers pass the Turing Test?"

The median estimate of the AI experts was for a 10% chance that computers will pass the Turing Test by 2022, rising to a 50% chance by 2040, and a 90% chance by 2075 (Baum et al., 2011). I predict that laypeople will overestimate, compared to AI experts, how long it will take for computers to pass the Turing Test.

Results

For the questions, "By what year would you see a 10%/50%/90% probability for such HLMI to exist?" the median responses of participants were years 2021, 2040, and 2055, respectively, compared to the median responses of experts, years 2022, 2045, and 2075, respectively. Thus, I did not find support for *Hypothesis 2*, that laypeople would overestimate, compared to AI experts, how long it will take for HLMI to exist.

For the questions, "By what year would you see a 10%/50%/90% probability that computers pass the Turing Test?" the median responses of participants were years 2020, 2030, and 2050, respectively, compared to the median responses of experts, years 2020, 2040, and 2075. Thus, I did not find support for *Hypothesis 2* that laypeople would overestimate compared to AI experts how long it will take for computers to pass the Turing Test.[3]

Study 3: Does the Exponential Growth Bias Become More Pronounced as Growth Rates Accelerate and Time Horizons Lengthen?

Study 3 tests whether participants become more inaccurate as growth rates increase (e.g., when the rate increases from 100% to 200%) and when time horizons lengthen (e.g., when time horizons lengthen from 20 years to 40 years). If participants become more inaccurate in each case, we might expect participants to become even more inaccurate in instances of rapid technological progress, and over longer periods of time.

Hypothesis 3: Participants become more inaccurate as growth rates increase and

[3] Inferential tests were not conducted as I do not have the data sets of the predictions made by experts.

as time horizons lengthen.

Sample

Four-hundred fifty-five participants in the United States completed the study via Amazon Mechanical Turk in exchange for $0.20 (60.0% female; mean age = 34.0).

Procedures

Participants answered one of six questions about exponential growth. Each question used a different exponential growth rate (i.e., 10%, 25%, 50%, 100%, 250%, 500%). For example, participants in the 10% condition answered the following question. “Imagine that a bank offers to give you $1 to deposit in their bank, and the bank agrees to give you 10% return on your money at the end of each year. For example, at the end of year 1, you would have $1.10; at the end of year 2 you would have $1.21; at the end of year 3 you would have $1.33, etc. How much money would you have: at the end of year 20? at the end of year 40? When answering the questions, please provide your thoughtful best guess. In other words, it is important that you do not formally calculate your answer (e.g., by using a calculator or scratch paper). We want your best guess.”

I predict that participants’ estimates will become progressively less accurate as growth rates and as time horizons lengthen.

Results

Performance scores were calculated by first calculating the absolute deviation for each of the items (i.e., the absolute value of the difference between participants’ answers and the correct answers; Laughlin & Ellis, 1986). Thus, lower scores indicate better accuracy. For example, a performance score of 0 means no deviation from the correct response.

As predicted, estimates tended to become progressively less accurate as growth rates and time horizons increased. In the 10% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 3.73 and 33.26. Moreover, 61.0% of participants underestimated the correct response at 20 years, $X^2(1, N = 77) = 3.75$, $p = .053$, and 79.2% of participants underestimated the correct response at 40 years, $X^2(1, N = 77) = 26.30$, $p < .001$.

In the 25% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 66 and 7,473. Moreover, 79.5% and 87.7% of participants underestimated the correct response at 20 years, $X^2(1, N = 73) = 25.33$, $p < .001$, and 40 years, $X^2(1, N = 73) = 41.44$, $p < .001$, respectively.

In the 50% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 3,285 and 11,057,197. Moreover, 97.6% and 100% of participants underestimated the correct response at 20 years, $X^2(1, N = 82) = 74.20$, $p < .001$ and 40 years (inferential test not conducted, given that 100% of participants underestimated the correct response at 40 years).

For the growth rates 100%, 250% and 500%, there appears to be a ceiling effect in terms of the percentage of participants who underestimated the correct answer regardless of the term. In the 100% growth rate condition, 92.3% and 94.9% of participants underestimated the correct response at 20 years, $X^2(1, N = 78) = 55.85$, $p < .001$, and 40 years, $X^2(1, N = 78) = 62.82$, $p < .001$, respectively. In the 250% growth rate condition, 98.6% of participants underestimated the correct response at both 20 years, $X^2(1, N = 72) = 68.06$, $p <.001$, and 40 years, $X^2(1, N = 72) = 68.06$, $p <.001$. In the 500% growth rate condition, 98.6% and 94.5% of participants underestimated the correct response at 20 years, $X^2(1, N = 73) = 69.06$, $p <.001$ and 40 years, $X^2(1, N = 73) = 57.88$, $p <.001$, respectively.

Participants' median performance scores increased as growth rates and time horizons increased. In the 100% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 1.03 million and 1.10 trillion. In the 250% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 76.10 billion and 5.79 sextillion. Finally, in the 500% growth rate condition, the median performance scores for years 20 and 40 were, respectively, 3.66 quadrillion and 13.37 nonillion. Thus, *Hypothesis 3* was supported.

Study 4: Do Participants Accept the Arguments, but Reject the Aversive Implications?

Study 4 examines whether participants agree with assumptions about technological growth but then reject the implications of those assumptions when the implications are revealed to be negative.

Hypothesis 4: Participants accept a given technological state and rate of technological progress, but then reject the implications of those assumptions when the implications are revealed to be aversive.

Sample

Eight-hundred two participants in the United States completed the study via Amazon Mechanical Turk in exchange for $0.10 (64.6% female; mean age = 35.7).

Procedure

Participants were assigned to one of two conditions (i.e., human, chimpanzee). The conditions were identical except for the word "human" was replaced with "chimpanzee" depending on condition. Participants in the human condition read the following:

> For the purpose of this question, "intelligence" is defined as *the ability to accomplish complex goals.* Choose the statement that you most agree with:
> Humans are one hundred times more intelligent than computers.
> Humans are one thousand times more intelligent than computers.
> Humans are one million times more intelligent than computers.
> Computers are one hundred times more intelligent than humans.
> Computers are one thousand times more intelligent than humans.
> Computers are one million times more intelligent than humans.

After participants made their selection they answered another question:

> Computers are becoming more intelligent each year. At what rate do computers become more intelligent each year?
> 1%
> 25%
> 50%
> 100%
> 250%

Thus, participants were allowed to a) choose the technological state of computer intelligence compared to human/chimpanzee intelligence, and b) choose the growth rate of computer progress. Once participants made their first two selections, they were shown the implications of their first two selections (i.e., their first two answers were calculated based on a time period of 50 years). For example, if a participant selected “Humans are one thousand times more intelligent than computers” and computers become “50%” more intelligent each year, the participant would then read the following:

> Based on your two prior responses ("humans are 1,000 times more intelligent than computers," and "computers become more intelligent at a rate of 50% per year") computers will be approximately 600,000 times more intelligent than humans in 50 years (assuming that humans do not become meaningfully more intelligent during this time period).

Once participants saw the implications (i.e., calculations) of their first two responses, participants then answered a question (i.e., the primary dependent variable) about the degree to which they disagreed/agreed (scale 1-7) with the implications of their respective responses: “How much do you agree with the statement that computers will be approximately 600,000 times more intelligent than humans in 50 years?”

Given humans’ desires to be competent and powerful (McClelland, 1965; Ryan & Deci, 2000), I predict participants will be less likely to agree with statements about computers surpassing human capabilities than computers surpassing chimpanzee capabilities. Or in other words, I predict that due to the motivated reasoning bias, participants in the human condition will be significantly less likely to agree with the statement than will participants in the chimpanzee condition, specifically when participants choose one of the final three growth rates (wherein the exponential growth bias would be exacerbated).

Results

Participants' ratings of how much they disagreed/agreed with the implications (i.e., calculations) of their first two responses was the dependent variable of interest. A between-groups ANOVA revealed a significant difference between the human (M = 4.29, SD = 1.63) and chimpanzee (M = 4.54, SD = 1.56) conditions, $F(1, 801) = 5.10$, $p = .024$, $\eta^2 = .01$. As predicted, participants in the human condition were significantly less likely to agree with the implications of their first two responses than were participants in the chimpanzee condition.

For participants who chose one of the latter three growth rates (N=318), I predicted that the difference between conditions would be even more pronounced. As predicted, a between-groups ANOVA revealed a significant difference between the human (M = 4.31, SD = 1.69) and chimpanzee (M = 4.89, SD = 1.54) conditions, $F(1, 317) = 10.09$, $p = .002$, $\eta^2 = .03$, and the effect was even more pronounced than when all participants were included. Thus, *Hypothesis 4* was supported.

Study 5: Do Participants Reject the Aversive and Accept the Desirable Implications of Exponential Technological Growth?

Study 5 examines whether participants tend to accept the positive implications of exponential technological growth, while rejecting the negative implications.

Hypothesis 5: Participants reject the aversive and accept the desirable implications of rapid technological progress.

Sample

Six-hundred thirty-nine participants in the United States completed the study via Amazon Mechanical Turk in exchange for $0.30 (59.2% female; mean age = 36.0).

Procedures

Participants were assigned to one of three conditions (i.e., control, positive, negative) and read four randomized statements about technological progress. After reading each statement, participants indicated (on a scale of 1-7) how much they disagreed/agreed with each of the four statements. Before answering each question, however, participants wrote one paragraph indicating how they would feel if the statement were true. Writing these paragraphs was intended to strengthen the manipulation. A complete list of the statements is provided in the Appendix. For example, in the control condition participants read the following: "By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today." In the positive condition participants read the following: "By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today, thus providing humans the freedom and autonomy to pursue their passions." In the negative condition participants read the following: "By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today, thus creating mass unemployment." The degree to which participants disagreed/agreed with each statement was the dependent variable of interest.

After participants rated each of the four statements (e.g., participants in the control condition always read statements from the control condition, although statements were randomized within condition), participants answered questions that assessed their general level of optimism, allowing me to control for trait optimism.

I predict that participants in the control condition and the positive condition will be more likely to agree with the statements than will participants in the negative condition.

Results

Participants' ratings of how much they disagreed/agreed with each of the four statements were highly correlated within condition (control condition α=.67; positive condition α = .71; negative condition α = .67), and thus were averaged to create a composite measure of disagree/agree, the dependent variable of interest. A one way, between-subjects ANOVA was conducted to compare the effect of condition on the primary dependent variable, while controlling for trait optimism.[4] Results indicate a main effect for condition, $F(2, 634) = 10.45$, $p < .001$, η^2 =.03. A Tukey post hoc test revealed that the negative condition (M = 3.66, SD = 1.25) significantly differed from both the control condition (M = 4.16, SD = 1.30) $p < .001$, d = *.39* and the positive condition (M = 4.15, SD = 1.31) $p < .001$, d = .38. As predicted, participants in the control and positive condition were significantly more likely to agree with the statements than participants in the negative condition. Thus, *Hypothesis 5* was supported.

---

[4] The effect of trait optimism on the dependent variable was not significant, p = .161.

Study 6: Does the Motivated Reasoning Bias Interact With the Exponential Growth Bias?

Study 6 examines whether the motivated reasoning bias interacts with the exponential growth bias.

Hypothesis 6a: Participants will be more likely to accept the positive versus negative consequences of technological progress.

Hypothesis 6b: Participants will be more likely to accept consequences resulting from a smaller growth rate than larger growth rate.

Hypothesis 6c: The exponential growth bias will be exacerbated for negative consequences compared to positive consequences.

Sample

Nine-hundred nine participants in the United States completed the study via Amazon Mechanical Turk in exchange for $0.30 (62.0% female; mean age = 37.1).

Procedures

Participants were randomly assigned to one of six conditions. The overall design of the study was a 2 (consequences: desirable vs. aversive) x 3 (growth rate: small vs. moderate vs. large) between-subjects design. Participants first read the following: "Take three minutes to write about your favorite leisure activity/hobby that you don't have enough time to pursue." After participants wrote for three minutes, they then read the following:

> Computers/robots are improving rapidly. According to some estimates, by 2018 computers/robots will [help humans complete (desirable) / replace

> humans in (aversive)] 50,000 of the 125 million jobs in the U.S. By year 2019, that number is estimated to rise by another [50,000 (small growth rate) / 10% (moderate growth rate) / 50% (large growth rate)], totaling [100,000 (small growth rate) / 55,000 (moderate growth rate) / 75,000 (large growth rate)] jobs. By year 2020, the number is estimated to rise by another [50,000 (small growth rate) / 60,500 (moderate growth rate) / 112,500 (large growth rate)] jobs. Thus, according to some estimates, the number of jobs that computers/robots will [help humans complete (desirable) / replace humans in (aversive)] is expected to rise by [50,000 (small growth rate) / 10% (moderate growth rate) / 50% (large growth rate)] each year for the foreseeable future.

Participants then read the following:

> Imagine a world where, due to advancements in technology (e.g.., [computers/robots helping humans in the workforce (desirable) / computers/robots replacing humans in the workforce (aversive))], you have [complete autonomy (desirable) / complete lack of autonomy (aversive)] to pursue the activities you most want to pursue (e.g., hobbies, work, travel, etc.). Please take three minutes to write about the [POSITIVE (desirable) / NEGATIVE (aversive)] consequences of such a world.

Participants then answered the following question:

> How much do you agree with the following statement: "By the year [4516 (small growth rate) / 2101 (moderate growth rate) / 2055 (large growth rate)] computers/robots will [help humans complete essentially all jobs, creating widespread abundance and prosperity, resulting in humans having the autonomy to pursue hobbies such as travel, art, health, music, sports, education, etc. (desirable) / replace humans in essentially all jobs, creating widespread unemployment and poverty, resulting in humans NOT having the autonomy to pursue hobbies such as travel, art, health, music, sports, education, etc. (aversive)]

I predict that the motivated reasoning bias and the exponential growth bias will interact, such that a) participants will be more likely to accept the positive versus negative consequences of technological progress; b) participants will be more likely to accept consequences resulting from a smaller growth rate than larger growth rate; and c) the exponential growth bias will be exacerbated for negative consequences compared to positive consequences.

Results

A 2 (consequence: desirable, aversive) x 3 (rate: small, moderate, large) between-subjects ANOVA was conducted to examine whether the motivated reasoning bias interacted with the exponential growth bias. The dependent variable for the test was the degree to which participants disagreed/agreed (scale 1-7) with the implications described in their respective conditions.

The effect of consequence was significant $F(2, 903)=26.85$, $p < .001$, $\eta^2 = .03$. Planned contrasts revealed that participants in the desirable condition ($M = 4.09$, $SD = 1.72$) were significantly more likely to accept the consequences of technological progress than were participants in the aversive condition ($M = 3.48$, $SD = 1.84$), $p < .001$, $d = .34$. Thus, *Hypothesis 6a* was supported

The effect of rate was also significant $F(2, 903) = 3.08$, $p = .047$, $\eta^2 = .01$. Planned contrasts revealed that participants in the small rate condition ($M = 3.99$, $SD = 1.85$) and the moderate rate condition ($M = 3.76$, $SD = 1.76$) did not significantly vary in the degree to which they accepted the consequences of technological progress. However, participants in the small rate condition were significantly more likely to accept the consequences of technological progress than were participants in the large rate condition ($M = 3.63$, $SD = 1.79$), $p = .038$, $d = .20$. Thus, *Hypothesis 6b* was supported.

Finally, the interaction between consequence and rate was not significant, $F(2, 903) = 1.27$, $p = .282$, $\eta^2 = .003$. It is possible that Study 6 was underpowered, and that with a larger sample, there would indeed be a significant interaction between consequence and rate. However, in the current form, *Hypothesis 6c* was not supported.

CHAPTER 3

GENERAL DISCUSSION

In this dissertation, I examined whether people are prone to underestimate technological progress, due to the exponential growth bias. I also examined whether people are prone to discount the aversive implications of rapid technological progress due to the motivated reasoning bias. Across six studies I found support for my hypotheses, to varying degrees.

In Study 1, participants underestimated the progress that AIs/robots have already made. For the tasks that AIs/robots have already completed, the majority of participants predicted that AIs/robots would complete the tasks either several years in the future or never at all (5% - 15% of participants predicted AIs/robots would never be able to complete the respective tasks). For the tasks that AIs/robots may have already completed, 15% - 30% of participants predicted that AIs would never be able to complete the respective tasks. Study 3 provided support for why participants would be prone to underestimate AI advancements. In Study 3, participants consistently, and vastly underestimated exponential growth curves, especially in instances of *larger* growth rates (e.g., 100%, 250%, and 500%) and *longer* time horizons (e.g., 40 years). Given that computing technology has progressed at a *large* exponential rate (i.e., 100% approximately every two years; Brynjolfsson & McAfee, 2014) for a *long* time period (i.e., approximately the last 50 years; Brynjolfsson & McAfee, 2014), I expected participants to underestimate the progress of AIs.

In Study 4, I found support for the hypothesis that participants engage in motivated reasoning when thinking about technological progress. Participants were more likely to accept the implications of rapid technological progress in the instance of computers surpassing chimpanzee ability, compared to computers surpassing human ability. Presumably it is more aversive (to humans) for computers to surpass human ability than for computers to surpass chimpanzee ability. Studies 5 and 6 provided additional support for the argument that participants are motivated to disagree with aversive implications of rapid technological progress. In studies 5 and 6, participants were more likely to accept the desirable implications of technological progress than the aversive implications of technological progress.

Not all predictions were supported, however. In Study 2, participants underestimated, compared to AI experts, how long it would take for computers to reach both human level intelligence and pass the Turing Test. In Study 6, participants were more likely to accept positive versus negative consequences of technological progress and to accept consequences resulting from smaller rather than larger growth rates as predicted. However, the two conditions (positive consequence vs. negative consequence and smaller growth rate vs. larger growth rate) did not interact as predicted.

Theoretical Contributions

The findings in the present research contribute to the literatures on perception and bias. Humans make predictable errors in judgment in numerous domains, but the exponential growth bias appears to be one of the most pronounced biases. This bias is understandable, given that few things in our daily experiences progress at an exponential rate. Thus, when we are confronted with exponential growth curves, the research indicates that we tend to underestimate our extrapolations (de Langhe et al., 2017). The current research extends our understanding exponential growth bias, and examines the phenomena in the domain of technological progress. Past work on the exponential growth bias has examined the phenomena in contexts such as forecasting pollution (Wagenaar & Sagaria, 1975) or financial returns (McKenzie & Liersch, 2011. To my knowledge, the current research is the first to examine the exponential growth bias in the context of technological progress.

Whereas past research documents the robustness of the exponential growth bias, the current research extends this work by also examining the magnitude of the bias. Exponential growth rates of 10% led to significant underestimation: 61.0% of participants, and 79.2% of participants underestimated the growth curves at terms of 20 and 40 years, respectively. Importantly, growth rates in computing power have been significantly faster than 10% over the last 40 years.

Whereas participants underestimate growth rates of 10%, participants especially underestimated larger growth rates. As growth rates increased to 50% and even 100% (e.g., the growth rates that approximate the improvements in computing power each year), more than 90% of participants underestimated the growth. As staggering as it is that 90% of participants underestimated the growth curves, it is also worth noting the magnitude of the underestimation. For a growth rate of 100% and a term of 20, the average underestimation was by 203% (e.g., if the correct answer were 100, the average estimate would have been 33); for a term of 40 years, the average underestimation was by 1,487% (e.g., if the correct answer were 100, the average estimate would have been just 6.3). As growth rates increased to 250% and 500%, the magnitude of the underestimation was greater than 4,248%, regardless of term.

Given the percentage of participants who underestimated exponential growth curves, and the magnitude of their underestimations, it is no surprise that the majority of participants in Study 1 underestimated current AI capabilities. Moreover, if people are unaware of current AI capabilities, it is reasonable to expect them to be surprised by future capabilities. For example, laypeople may be surprised by how soon self-driving cars become prevalent on roadways. At a recent conference, I asked Stuart Russell—the author of the most widely cited textbook on AI— when he expected self-driving cars to drive 90% of the miles in the United States. Stuart responded, “I am more pessimistic than most experts. I think it probably won’t happen until 2025 or 2030.” In my experience, laypeople do not expect self-driving cars to be so prevalent in the next decade. For example, in a pilot study I recently ran, I asked 100 people on Amazon Mechanical Turk to predict when self-driving cars would drive 90% of the miles in the United States. The median response of participants was year 2055.

Not only are participants prone to underestimate technological progress due to the exponential growth bias, but also the current research suggests that participants are prone to erroneous expectations due to the motivated reasoning bias. People are prone to believe in a favorable future in terms of politics, scientific beliefs, and entertainment preferences (Rogers et al., 2017). As a corollary, the current research suggests that people are also prone to disbelieve an unfavorable future in terms of technological progress.

Participants in Studies 4, 5, and 6 were significantly more likely to accept the implications of technological progress when the implications were neutral or positive compared to when the implications were negative. However, technological progress may indeed lead to numerous negative outcomes, and people's desire for a favorable future may impede them from accurately assessing current and/or future technological capabilities.

The current research on motivated reasoning is further bolstered by a recent Gallup survey which found that most Americans think AI will improve lives and eliminate jobs — just not their jobs (Carlson, 2018). Moreover, 75% of respondents said they are optimistic about the impact AI will have on their lives, and yet nearly the same proportion said that AI will result in a net loss of jobs. Thus, the current research provides additional evidence to the growing body of research that suggests that people are motivated to draw conclusions that are desirable, especially in terms of technological progress.

Possible Limitations

The studies presented herein are limited in several ways. A primary objective of this dissertation is to examine whether people are ultimately unprepared for the implications of rapid technological progress. To answer this question, I examine underlying psychological phenomena in an attempt to assess people's awareness/readiness. However, longitudinal tests that compare people's predictions of AI advancements with the eventual, actual occurrences of progress could provide additional evidence regarding tendencies to over/underestimate.

In Study 1, I ask participants to make estimates about 12 tasks. Although participants underestimated AI progress on the four tasks that AIs have already completed, it is possible that participants are accurately calibrated on other tasks that were not presented. In sum, Study 1 examines a narrow set of tasks that may not be representative of all the relevant tasks that people could make predictions about.

In Study 2, I compare participants' responses with the responses of AI experts. It is possible, however, that the experts' predictions are less accurate than the laypeople's predictions. For example, it may be the case that experts, as a collective, tend to overestimate how long it will take for AIs to pass the Turing Test. The experts presumably have a much closer relationship with AI technology than laypeople, and thus may be more sensitive to the magnitude of the innovation required for computers to pass the Turing Test. This enhanced familiarity with the technical challenges required may lead experts to focus more on the challenges present, rather than the innovations that are possible, thus biasing their responses. Alternatively, the experts may be overly optimistic about their estimates due to the motivated reasoning bias, as presumably many of the experts are working on solving the problems related to AIs passing the Turing Test. Finally, the experts responded to these questions between the years 2009-2013, whereas the participants in Study 2 responded in the year 2017. Thus, it is possible that if experts were asked the same questions today, their predictions would be even more optimistic/pessimistic than they were several years ago.

In Study 3, I asked participants to make their estimates based on their intuitions, and specifically asked participants to not calculate their answers. The purpose for these instructions was to get a sense for participants' intuitions around exponential growth curves, rather than assess whether participants could actually calculate the answers. However, in the case of technological progress, it is possible for people to actually run the calculations. People can examine how quickly computing speed is accelerating (and has accelerated) and then calculate where computing speed may be in the near and long term (assuming continuous growth rates). In sum, a weakness of Study 3, as it relates to the broader question of whether people are underestimating AI, is that people can run the calculations, given a set of assumptions. People will not be left to just their intuitions when making assessments of technological progress.

In Study 4, I ask participants to make assessments about the intelligence of humans/chimpanzees and computers. It is possible that participants do not agree that the intelligence of humans/chimpanzees is on the same plane or even dimension as the intelligence of computers, and thus participants' answers may have been skewed in unforeseen ways.

In Study 5, I ask participants to make assessments about the degree to which they disagree/agree with statements about technological progress. But similar to the weaknesses described in Study 1, Study 5 only uses four questions to assess participant's response to the positive/negative consequences of AI. It is possible that if participants answered questions about different tasks, that their responses would change. In sum, Study 5 examines a narrow set of tasks that may not be representative of all the relevant tasks that people could disagree/agree with.

In Study 6, I found no interaction between consequences and growth rates as predicted. However, it may be the case that an interaction would have been found had the sample size been larger. It is possible that I overestimated the effect size when calculating sample size. It does seem logical and even plausible that the exponential growth bias would be exacerbated for negative consequences compared to positive consequences. Although I did not find support for the hypothesis, it does not preclude the possibility that the predicted relationship exists, as absence of evidence is not evidence of absence.

Future Directions

For future directions, I plan to examine additional moderators of the effects found herein. For example, I plan to examine whether people's personalities or political orientations are predictive of their perceptions of AI. Future research could also shed additional light on the moderators of motivated reasoning. Is there a particular personality type that is more prone to motivated reasoning than others when it comes to technological progress?

I also plan to examine people's perceptions of AI adoption. For example, I am starting a project where I plan to collect data from doctors who are encouraged to adopt algorithmic diagnoses in their practices. I want to better understand the factors that predict whether people will trust and incorporate the recommendations of AIs/algorithms.

I also plan to examine numerous ethical questions related to AI adoption. For example, are people more or less ethical when making decisions in conjunction with AIs? Employees of Lyft and Uber continuously interact with algorithms (e.g., the algorithm indicates where the drivers should pick up and carry passengers). Does working with AIs lead people to dehumanize others? Does working with AIs make people happy? I am also intrigued by whether people are more likely to engage in unethical behavior if AIs recommend an unethical course of action, compared to when humans suggest an unethical course of action.

Finally, I hope to examine people's perceptions of job satisfaction as AIs become more prevalent in the work force. For example, do employees feel like they have more or less voice as AIs continue to become more prevalent. And if AIs do indeed replace vast numbers of human employees, I question whether those humans who are out of work will be able to find meaning in life.

Practical Implications

The research herein indicates that people may indeed be both unaware of the consequences, and unaccepting of the implications, of rapid technological progress. The current research, while conducted at a psychological level, has implications for organizations and societies at large.

At an organizational level, the capability of AI has the potential to outpace managers' expectations for it—especially if growth rates accelerate. Managers wanting to take advantage of a technology that does not currently exist, may be prone to overestimate how long it will take for the technology to be developed (due to the exponential growth bias). Thus, managers would be prone to adopt the use of AI later than is optimal when adoption requires lead-time. Forward-thinking managers would start making preparations for technological adoptions sooner than their intuitions may lead them to expect. An ability to forecast AI progress (i.e., exponential growth) would thus be required of those decision makers aiming to adopt AI at the optimal time. Moreover, given the accelerating rates of product adoption (Dobbs et al., 2015), and thus the enhanced competitive advantage enjoyed by first-movers, the importance of accurately assessing AI capabilities is hard to overstate.

Societies at large may also find themselves unprepared for the changes that are coming if computing technology (generally) and AI (specifically) continue to progress at a rapid rate. Raising awareness of the potential for rapid change is a logical first step in preparing people and organizations. The cost of a false positive (making people aware of the potential for drastic change that never actually happens) seems to pale in comparison to the cost of the false negative (people remaining unaware of the drastic changes until it is too late).

The current research also has implications for policy makers as they address issues such as technological unemployment (i.e., the loss of jobs to due technological change). If indeed AIs/robots continue to progress at a rapid rate, it will not be long until they are capable of completing most tasks that humans can complete today, potentially causing massive unemployment. The current research suggests that we may be prone to underestimate how quickly we may need to prepare for a workplace that is dominated by AIs/robots. Thus, I hope organizational and governmental leaders pay close attention to AI progress.

Conclusion

In this dissertation, I have examined one of the most (potentially) pressing issues of our time. The rapid advancement of technology has the potential to fundamentally alter our lives and the organizations we are part of. If technology continues to progress at an exponential rate, we are prone to vastly underestimate its impact. Moreover, due to the motivated reasoning bias, our expectations may prove to be out of sync with reality. An awareness of these biases and their implications is the first step in preparing for a future that may be vastly different from our present.

APPENDIX

Statement 1:

Control: By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today.

Positive: By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today, thus providing humans the freedom and autonomy to pursue their passions.

Negative: By the year 2060, computers/robots will be capable of completing most tasks that humans can complete today, thus creating mass unemployment.

Statement 2:

Control: By 2040, the only cars on the road will be self-driving cars.

Positive: By 2040, the only cars on the road will be self-driving cars, allowing people to pursue enjoyable activities (e.g., sleeping, eating, watching movies) while in their cars.

Negative: By 2040, the only cars on the road will be self-driving cars, and thus humans will no longer be permitted to drive.

Statement 3:

Control: By 2080 computers will be far more capable than humans in every way.

Positive: By 2080 computers will be far more capable than humans in every way, allowing computers to solve problems such as world hunger.

Negative: By 2080 computers will be far more capable than humans in every way, allowing computers to govern humans.

Statement 4:

Control: By 2080, robots will develop meaningful relationships with humans.

Positive: By 2080, robots will develop meaningful relationships with humans, and serve as companions to people who are lonely.

Negative: By 2080, robots will develop meaningful relationships with humans, and thus robot friends will replace human friends to some degree.

REFERENCES

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, *11*(1), 109–126. http://doi.org/10.1017/CBO9781107415324.004

Balog, M., Gaunt, A. L., Brockschmidt, M., Nowozin, S., & Tarlow, D. (2017). DeepCoder: Learning to write programs. In *International Conference on Learning Representations*.

Baum, S. D., Goertzel, B., & Goertzel, T. G. (2011). How long until human-level AI? Results from an expert assessment. *Technological Forecasting and Social Change*, *78*(1), 185–195.

Bazerman, M. H., & Moore, D. A. (2013). *Judgment in managerial decision making (8th Edition)*. New York, NY: Wiley.

Bhuiyan, J. (2016). After two million miles, Google's robot car now drives better than a 16-year-old. Retrieved February 8, 2017, from http://www.recode.net/2016/10/5/13167364/google-self-driving-cars-2-million-miles

Blascovich, J., Loomis, J., Bell, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology: Just another methodological tool for social psychology? *Psychological Inquiry*, *13*(2), 146–149.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, and strategies*. Oxford, UK: Oxford University Press.

Bostrom, N., & Müller, V. (2014). Future progress in artificial intelligence: A survey of expert opinion. In V. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 553–570). Berlin, Germany: Springer.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* New York, NY: WW Norton & Company.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247.

Buehler, R., Messervey, D., & Griffin, D. W. (2005). Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation. *Organizational Behavior and Human Decision Processes*, *97*(1), 47–63.

Burrus, D. (2017). The Internet of things is far bigger than anyone realizes. Retrieved February 14, 2017, from https://www.wired.com/insights/2014/11/the-internet-of-things-bigger/

Cadwalladr, C. (2014). Are the robots about to rise? Google's new director of engineering thinks so…. Retrieved February 13, 2017, from https://www.theguardian.com/technology/2014/feb/22/robots-google-ray-kurzweil-terminator-singularity-artificial-intelligence

Carlson, S. (2018). Most Americans think AI will improve lives and eliminate jobs — just not theirs. Retrieved January 31, 2018, from https://www.chronicle.com/article/Most-Americans-Think-AI-Will/242386?cid=wcontentlist_hp_latest

Chou, M., Han, S. H., Lu, B., Yoshida, Y., Taguchi, H., Kataura, T., … Lin, F. (2017). *AI - New computing revolution*. Hong Kong, China: Deutsche Bank Markets Research.

Courtland, R. (2017). Intel now packs 100 million transistors in each square millimeter. Retrieved March 30, 2017, from http://spectrum.ieee.org/nanoclast/semiconductors/processors/intel-now-packs-100-million-transistors-in-each-square-millimeter

Davidson, S. (2106). Artificial intelligence comes to mortgages but would YOU feel comfortable taking financial advice from a machine? Retrieved February 1, 2017, from http://www.thisismoney.co.uk/money/mortgageshome/article-3765065/Artificial-intelligence-comes-mortgages-feel-comfortable-machine-giving-financial-advice.html

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin*, *28*(10), 1379–1387. http://doi.org/10.1177/014616702236869

de Langhe, B., Puntoni, S., & Larrick, R. (2017). Linear thinking in a nonlinear world. *Harvard Business Review*, (May), 130–139.

Derry, T. K., & Williams, T. I. (1993). *A short history of technology: From earliest times to A.D. 1900*. Mineola, NY: Dover.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., … Morency, L. (2014). SimSensei Kiosk : A virtual human interviewer for healthcare decision support. *International Conference on Autonomous Agents and Multi-Agent Systems*, 1061–1068.

Dobbs, R., Manyika, J., & Woetzel, J. (2015). *No ordinary disruption*. New York, NY: PublicAffairs.

Dormehl, L. (2017). *Thinking machines: The quest for artificial intelligence and where it's taking us next*. New York, NY: TarcherPerigee.

Dryer, D. C., & Horowitz, L. M. (1997). When do opposites attract? Interpersonal complementarity versus similarity. *Journal of Personality and Social Psychology*, *72*(3), 592–603. http://doi.org/10.1037/0022-3514.72.3.592

Eck, D. (2016). Welcome to Magenta! Retrieved February 9, 2017, from https://magenta.tensorflow.org/welcome-to-magenta

Elon Musk enters the world of brain-computer interfaces. (2017, March). Retrieved April 3, 2017, from http://www.economist.com/news/science-and-technology/21719774-do-human-beings-need-embrace-brain-implants-stay-relevant-elon-musk-enters

Evans, R., & Gao, J. (2016). DeepMind AI reduces Google data centre cooling bill by 40%. Retrieved February 1, 2017, from https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

Feldman, M. (2016). Market for artificial intelligence projected to hit $36 billion by 2025. Retrieved January 17, 2017, from www.top500.org/news/market-for-artificial-intelligence-projected-to-hit-36-billion-by-2025/

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: beyond Jeopardy!. *Artificial Intelligence*, *199*, 93–105.

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. New York, NY: Basic Books.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. http://doi.org/10.1016/j.techfore.2016.08.019

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814–821. http://doi.org/10.1006/nimg.2002.1117

Garber, M. (2014). Would you want therapy from a computerized psychologist? Retrieved February 8, 2017, from https://www.theatlantic.com/technology/archive/2014/05/would-you-want-therapy-from-a-computerized-psychologist/371552/

Gibbs, S. (2016). Chatbot lawyer overturns 160,000 parking tickets in London and New York. Retrieved February 13, 2017, from https://www.theguardian.com/technology/2016/jun/28/chatbot-ai-lawyer-donotpay-parking-tickets-london-new-york

Gilovich, T. (1983). Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology*, *44*(6), 1110–1126. http://doi.org/10.1037/0022-

3514.44.6.1110

Goldhill, O. (2016). The first pop song ever written by artificial intelligence is pretty good, actually. Retrieved February 1, 2017, from https://qz.com/790523/daddys-car-the-first-song-ever-written-by-artificial-intelligence-is-actually-pretty-good/

Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. *Advances in Computers*, *6*, 31–88. http://doi.org/10.1016/S0065-2458(08)60418-0

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Graham, L. (2017). Finland experiments with universal basic income scheme. Retrieved February 1, 2017, from http://www.cnbc.com/2017/01/03/finland-experiments-universal-basic-income.html

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*, 619. http://doi.org/10.1126/science.1134475

Gregory, L. (2016). A year on from winning TechCrunch Disrupt. Retrieved February 9, 2016, from http://blog.jukedeck.com/

Hal, 90210. (2016). This is what happens when an AI-written screenplay is made into a film. Retrieved February 9, 2017, from https://www.theguardian.com/technology/2016/jun/10/artificial-intelligence-screenplay-sunspring-silicon-valley-thomas-middleditch-ai

Hernandez, D., & King, R. (2016, November 24). Universities' AI talent poached by tech giants. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/universities-ai-talent-poached-by-tech-giants-1479999601

Hof, R. D. (2013). Deep learning. Retrieved February 7, 2017, from https://www.technologyreview.com/s/513696/deep-learning/

Hubbard, E., Matthews, P., & Samek, A. (2016). Using online compound interest tools to improve financial literacy. *The Journal of Economic Education*, *47*(2), 106–120. http://doi.org/10.1080/00220485.2016.1146097

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 482–526). New York, NY: Guilford.

Jones, D. A. (1991). Electrical engineering: The backbone of society. *IEE Proceedings: Science, Measurement and Technology*, *138*, 1–10.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–60. http://doi.org/10.1126/science.aaa8415

Kaku, M. (2012). *Physics of the future: How science will shape human destiny and our daily lives by the year 2100.* New York, NY: Doubleday.

Kappes, A., Faber, N. S., Kahane, G., Savulescu, J., & Crockett, M. J. (2018). Concern for others leads to vicarious optimism. *Psychological Science*, 1–27. http://doi.org/10.1177/0956797617737129

Katz, D. M., Bommarito II, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the supreme court of the United States. *PLoS One*, *12*(4), 1–18. http://doi.org/https://doi.org/10.1371/journal.pone.0174698

Khan, A. (2016, September 1). Could artificial intelligence help humanity? Two California universities think so. *L.A. Times*. Retrieved from http://www.latimes.com/science/sciencenow/la-sci-sn-usc-ai-social-good-20160831-snap-story.html

Khatchadourian, R. (2015, November). The dommsday invention. *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom

Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., … Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters*, *454*(3), 176–181. http://doi.org/10.1016/j.neulet.2009.03.026

Kitroeff, N. (2016, September 25). Robots could replace 1.7 million American truckers in the next decade. *L.A. Times*. Retrieved from http://www.latimes.com/projects/la-fi-automated-trucks-labor-20160924/

Knight, W. (2016). An AI ophthalmologist shows how machine learning may transform medicine. Retrieved February 1, 2017, from https://www.technologyreview.com/s/602958/an-ai-ophthalmologist-shows-how-machine-learning-may-transform-medicine/

Kopetz, H. (2011). *Internet of things*. *Real-time systems*. New York, NY: Springer.

Korn, M. (2016). Imagine discovering that your teaching assistant really is a robot. Retrieved January 17, 2017, from http://www.wsj.com/articles/if-your-teacher-sounds-like-a-robot-you-might-be-on-to-something-1462546621

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE*, *3*(7). http://doi.org/10.1371/journal.pone.0002597

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647. http://doi.org/10.1037/0022-3514.53.4.636

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kurzweil, R. (2005). *The singularity is near*. New York, NY: Penguin Books.

Lafferty, J. C., & Eady, P. . . (1974). *The desert survival problem*. Plymouth, MI: Experimental Learning Methods.

Laughlin, P. (1980). Social combination processes of cooperative problem-solving groups on verbal intellective tasks. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 127-155). Hillsdale, NJ: Erlbaum.

Laughlin, P., & Ellis, A. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology*, 177–189.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. http://doi.org/10.1038/nature14539

Levandowski, A. (2016). Pittsburgh, your self-driving uber is arriving now. Retrieved January 17, 2017, from https://newsroom.uber.com/pittsburgh-self-driving-uber

Lewis-Kraus, G. (2016). The great A.I. awakening. Retrieved February 20, 2017, from https://mobile.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html?_r=0&referer=https://t.co/aOEMZ7ozUZ

Liebermann, M. V., & Montgomery, D. B. (1988). First-mover advantages. *Strategic Management Journal*, *9*, 41–58. http://doi.org/10.1002/smj.4250090706

Littlepage, G. E., Robison, W., & Reddington, K. (1997). Effects of task experience and group experience on group performance, member ability, and recognition of expertise. *Organizational Behavior and Human Decision Processes*, *69*, 133–147. http://doi.org/10.1006/obhd.1997.2677

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, *69*(5), 877–889. http://doi.org/10.1037/0022-3514.69.5.877

Long, E., Lin, H., Liu, Z., Wu, X., Wang, L., Jiang, J., … Liu, Y. (2017). An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nature Biomedical Engineering*, *1*(January), 1–8. http://doi.org/10.1038/s41551-016-0024

Markoff, J. (2016). New research center to explore ethics of artificial intelligence. *New York Times*. Retrieved from https://www.nytimes.com/2016/11/02/technology/new-research-center-to-explore-ethics-of-artificial-intelligence.html?_r=4&mtrref=undefined

Marshall, A. (2017, January 22). From jingles to pop hits, A.I. is music to some ears. *New York Times*. Retrieved from https://www.nytimes.com/2017/01/22/arts/music/jukedeck-artificial-intelligence-songwriting.html?smid=tw-share&_r=0

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 11832–11835. http://doi.org/10.1073/pnas.211415698

McClelland, D. C. (1965). Toward a theory of motive acquisition. *American Psychologist*, *20*(5), 321–333. http://doi.org/10.1037/h0022225

McKenzie, C. R. ., & Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research (JMR)*, *48*(2011), S1–S13. http://doi.org/10.1509/jmkr.48.SPL.S1

McLean, A. (2017). Carnegie Mellon artificial intelligence victorious in heads-up poker tournament. Retrieved February 1, 2017, from http://www.zdnet.com/article/carnegie-melon-artificial-intelligence-victorious-in-heads-up-poker-tournament/

Melo, C. De, Marsella, S., & Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(2), 8. http://doi.org/10.1145/2890495

Metz, C. (2016). The rise of the artificially intelligent hedge fund. Retrieved February 1, 2017, from https://www.wired.com/2016/01/the-rise-of-the-artificially-intelligent-hedge-fund/

Microsoft. (2009). Video games and artificial intelligence. Retrieved February 13, 2017, from https://www.microsoft.com/en-us/research/project/video-games-and-artificial-intelligence/

Motor vehicle deaths increase by largest percent in 50 years. (2016). Retrieved February 8, 2017, from http://www.nsc.org/Connect/NSCNewsReleases/Lists/Posts/Post.aspx?List=1f2e4535-5dc3-45d6-b190-9b49c7229931&ID=103&var=hppress&Web=36d1832e-7bc3-4029-98a1-317c5cd5c625

Moukheiber, Z. (2014). Kaiser permanente chairman Bernard Tyson on the “Kaiserfication” of America. Retrieved February 8, 2017, from https://www.forbes.com/sites/zinamoukheiber/2014/09/30/kaiser-permanente-chairman-bernard-tyson-on-the-kaiserfication-of-america/#7ca6402b5948

Mullen, E., & Skitka, L. J. (2006). Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *Journal of Personality and Social Psychology*, *90*(4), 629–643. http://doi.org/10.1037/0022-

3514.90.4.629

Muoio, D. (2016). Washington D. C. residents can ride in this adorable driverless shuttle starting this summer. Retrieved March 15, 2017, from http://www.businessinsider.com/olli-driverless-shuttle-hits-dc-roads-2016-6

Murphy, J. (2017). Partnership on AI update. Retrieved January 31, 2017, from https://www.partnershiponai.org/2016/09/industry-leaders-establish-partnership-on-ai-best-practices/

Ng, A. (2016). Hiring your first chief AI officer. *Harvard Business Review*. Retrieved from https://hbr.org/2016/11/hiring-your-first-chief-ai-officer

Occupational employment statistics. (2015). Retrieved March 9, 2017, from https://www.bls.gov/oes/current/oes434051.htm

Oremus, W. (2016). The AP's new baseball reporter is not human. Retrieved February 6, 2017, from http://www.slate.com/blogs/future_tense/2016/06/30/the_ap_s_new_minor_league_baseball_reporter_is_a_robot_journalist.html

Overly, S. (2017, February 1). The great artificial intelligence gamble that finally paid off. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/innovations/wp/2017/02/01/the-great-artificial-intelligence-gamble-that-finally-paid-off/?utm_term=.ae589b2fff69

Paharia, N., Vohs, K. D., & Deshpandé, R. (2013). Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior and Human Decision Processes*, *121*(1), 81–88. http://doi.org/10.1016/j.obhdp.2013.01.001

Panko, B. (2017). Can eagle-eyed artificial intelligence help prevent children from going blind? Retrieved February 1, 2017, from http://www.smithsonianmag.com/science-nature/eagle-eyed-ai-doctor-could-nip-cataracts-bud-180961993/

Peltz, J. F. (2017, January 16). When Trump tweets, Wall Street trades — instantly. *L.A. Times*.

Perloff, L. S. (1983). Perceptions of vulnerability to victimization. *Journal of Social Issues*, *3*(2), 41–61.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205. http://doi.org/10.1558/ijsll.v14i2.309

Phillips, M. (2013). How the robots lost: High-frequency trading's rise and fall. Retrieved February 9, 2017, from https://www.bloomberg.com/news/articles/2013-06-06/how-the-robots-lost-high-frequency-tradings-rise-and-fall

Porter, M. E. (1985). Technology and competitive advantage. *Journal of Business Strategy*, *5*(3), 60–78. http://doi.org/10.1108/eb039075

Porter, M. E., & Millar, V. E. (1985). How information gives you competitive advantage. *Harvard Business Review*, *63*(4), 149–152. http://doi.org/10.1038/bdj.2007.481

Press, A. (2016). AP expands minor league baseball coverage. Retrieved February 8, 2017, from https://www.ap.org/press-releases/2016/ap-expands-minor-league-baseball-coverage

Ramsey, M., & Macmillan, D. (2015, May 31). Carnegie Mellon reels after Uber lures away researchers. *Wall Street Journal*. Retrieved from https://www.wsj.com/article_email/is-uber-a-friend-or-foe-of-carnegie-mellon-in-robotics-1433084582-lMyQjAxMTE1MjA5MTUwNzE5Wj

Randazzo, S. (2016). Data tools offer hints at how judges might rule. Retrieved February 1, 2017, from https://www.wsj.com/articles/data-tools-offer-hints-at-how-judges-might-rule-1481641205

Rayess, R. (2015). Is the CIO the next VP of electricity? Retrieved March 9, 2017, from https://techcrunch.com/2015/10/04/is-the-cio-the-next-vp-of-electricity/

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405. http://doi.org/10.1016/S0896-6273(02)00755-9

Roberts, S. (2016). The best or worst thing to happen to humanity. Retrieved January 31, 2017, from http://www.cam.ac.uk/research/news/the-best-or-worst-thing-to-happen-to-humanity-stephen-hawking-launches-centre-for-the-future-of

Robotics-VO. (2013). *A roadmap for us robotics: From internet to robotics*. *Computing Community Consortium*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Roadmap+for+U.S.+Robotics:+From+Internet+to+Robotics#0

Rogers, T., Moore, D. A., & Norton, M. I. (2017). The belief in a favorable future. *Psychological Science*, *28*(9), 1290–1301. http://doi.org/10.1177/0956797617706706

Rose, C. (2016). A.I. making a difference in cancer care. Retrieved February 20, 2017, from http://www.cbsnews.com/news/artificial-intelligence-making-a-difference-in-cancer-care/

Rosenbaum, E. (2016). Can elite law firms survive the rise of artificial intelligence? The jury is still out. Retrieved February 1, 2017, from http://www.cnbc.com/2016/11/17/can-cash-cow-of-elite-legal-firms-survive-ai-the-jury-is-still-out.html

Rosoff, M. (2014). Your phone is more powerful than the computer in the spaceship NASA launched this week. Retrieved February 6, 2017, from http://www.businessinsider.com/your-phone-is-more-powerful-than-the-orion-computer-2014-12

Rousseau, D. M., & Tijoriwala, S. a. (1999). What's a good reason to change? Motivated reasoning and social accounts in promoting organizational change. *Journal of Applied Psychology*, *84*(4), 514–528. http://doi.org/10.1037/0021-9010.84.4.514

Ruger, T. W., Kim, P. T., Martin, A. D., & Quinn, K. M. (2004). The supreme court forecasting project: Legal and political science approaches to predicting supreme court decisionmaking. *Columbia Law Review*, *104*(4), 1150–1210.

Russel, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. http://doi.org/10.1037/0003-066X.55.1.68

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*(5626), 1755–1758. http://doi.org/10.1126/science.1082976

Schaller, R. R. (1997). Moore's law: Past, present and future. *IEEE Spectrum*, *34*, 52–59.

Schubarth, C. (2017). Merck taps Palantir for help in speeding up drug discoveries. Retrieved January 31, 2017, from http://www.bizjournals.com/sanjose/news/2017/01/12/merck-taps-palantir-for-help-in-speeding-up-drug.html

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. http://doi.org/10.1038/nature16961

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. http://doi.org/10.1177/0956797611417632

Sobowale, J. (2016). How artificial intelligence is transforming the legal profession. Retrieved February 13, 2017, from http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession

Soll, J. B., Keeney, R. L., & Larrick, R. P. (2013). Consumer misunderstanding of credit card use, payments, and debt: Causes and solutions. *Journal of Public Policy and Marketing*, *32*(1), 66–81.

Sony. (2016). AI makes pop music. Retrieved February 9, 2017, from http://www.flow-machines.com/ai-makes-pop-music/

Standage, T. (2016). The return of the machinery question. *The Economist*. Retrieved from http://www.economist.com/news/special-report/21700761-after-many-false-starts-artificial-intelligence-has-taken-will-it-cause-mass

Stango, V., & Zinman, J. (2009a). Exponential growth bias and household finance. *The Journal of Finance*, *64*(6), 2807–2849.

Stango, V., & Zinman, J. (2009b). Internet appendix for "Exponential growth bias and household finance." *Journal of Finance*, (64), 2807–2849. Retrieved from http://www.afajof.org/IA/2009.asp#dec09

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Teller, A. (2016). *Artificial intelligence and life in 2030. One hundred year study on artificial intelligence: Report of the 2015-2016 study panel*. Stanford, CA. Retrieved from http://ai100.stanford.edu/2016-report

Synnott, W. R., & Gruber, W. H. (1981). *Information resource management: Opportunities and strategies for the 1980s*. New York, NY: John Wiley & Sons, Inc.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research, 169*(4), 564–568. http://doi.org/10.1007/s00221-006-0346-5

van den Oord, A., Zen, H., & Dieleman, S. (2016). Wavenet: A generative model for raw audio. Retrieved January 17, 2017, from https://deepmind.com/blog/wavenet-generative-model-raw-audio/

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Nodell, M. (2001). The sequence of the human genome. *Science*, *291*, 1304–1351. http://doi.org/10.1126/science.1058040

Wagenaar, W. A., & Sagaria, S. (1975). Misperception of exponential growth. *Perception and Psychophysics*, *18*, 416–422.

Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Perception and Psychophysics*, *24*, 182–184.

Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem; Three experiments on the misperception of exponential growth. *Acta Psychologica*, *43*(3), 239–251. http://doi.org/10.1016/0001-6918(79)90028-3

Waldrop, M. M. (2015). Autonomous vehicles: No drivers required. *Nature*, *518*, 20–3.

http://doi.org/10.1038/518020a

Waldrop, M. M. (2016). More than Moore. *Nature*, *530*, 144–147.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*, 383–388. http://doi.org/10.1016/j.tics.2010.05.006

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*, 802–820.

Weinstein, N. D. (1983). Reducing unrealistic optimism about illness susceptibility. *Health Psychology*, *2*(1), 11–20. http://doi.org/10.1037/0278-6133.2.1.11

Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology*, *3*(5), 431–457. http://doi.org/10.1037/0278-6133.3.5.431

Whitler, K. (2016). How artificial intelligence is changing the retail experience for consumers. Retrieved February 23, 2017, from http://www.forbes.com/sites/kimberlywhitler/2016/12/01/how-artificial-intelligence-is-changing-the-retail-experience-for-consumers/#60c38a8c1d6a

Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2010). Internet of things. *International Journal of Communication Systems*, *23*(5), 633–652. http://doi.org/10.1002/dac

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality*, *47*(2), 245–287. http://doi.org/10.1111/j.1467-6494.1979.tb00202.x