Blurring the Line Between Human and Machine: Marketing Artificial Intelligence

Noah Castelo

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Blurring the Line Between Human and Machine: Marketing Artificial Intelligence

Noah Castelo

One of the most prominent and potentially transformative trends in society today is machines becoming more human-like, driven by progress in artificial intelligence. How this trend will impact individuals, private and public organizations, and society as a whole is still unknown, and depends largely on how individual consumers choose to adopt and use these technologies. This dissertation focuses on understanding how consumers perceive, adopt, and use technologies that blur the line between human and machine, with two primary goals. First, I build on psychological and philosophical theories of mind perception, anthropomorphism, and dehumanization, and on management research into technology adoption, in order to develop a theoretical understanding of the forces that shape consumer adoption of these technologies. Second, I develop practical marketing interventions that can be used to influence patterns of adoption according to the desired outcome.

This dissertation is organized as follows. Essay 1 develops a conceptual framework for understanding what AI is, what it can do, and what are some of the key antecedents and consequences of its' adoption. The subsequent two Essays test various parts of this framework. Essay 2 explores consumers' willingness to use algorithms to perform tasks normally done by humans, focusing specifically on how the nature of the task for which algorithms are used and the human-likeness of the algorithm itself impact consumers' use of the algorithm. Essay 3 focuses on the use of social robots in consumption contexts, specifically addressing the role of robots' physical and mental human-likeness in shaping consumers' comfort with and perceived usefulness of such robots.

Together, these three Essays offer an empirically supported conceptual structure for marketing researchers and practitioners to understand artificial intelligence and influence the processes through which consumers perceive and adopt it. Artificial intelligence has the potential to create enormous value for consumers, firms, and society, but also poses many profound challenges and risks. A better understanding of how this transformative technology is perceived and used can potentially help to maximize its potential value and minimize its risks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGEMENTS

First and foremost, thank you to my family for everything that you did to get me to the point of finishing a PhD. My mother Marisa, who always wanted to finish her own PhD but never had the chance, my grandparents Jose and Adelle, my father Jan, my aunts Kate and Anna, each of you have been unconditionally loving, supportive, encouraging, patient, and wise.

Thank you also to my dissertation committee: Bernd Schmitt, Miklos Sarvary, Don Lehmann, Kate White, and Adrian Ward. The fact that I found my way into marketing academia at all is thanks to Kate taking a chance on me when I emailed out of the blue her asking if she would supervise my master's degree and then mentoring and collaborating with me ever since. Bernd and Miklos were willing to study robots and cyborgs with me before it was fashionable and have been a source of constant advice, ideas, guidance, and fun. Working with Don has made me a better researcher in many ways, and his humility and easygoing approach have been a great influence as well. Adrian is one of the most rigorous and thoughtful researchers I've met and has been a great friend too. Maarten Bos has also been a wonderful collaborator and friend who contributed a great deal to Essay 2 of this dissertation.

Finally, thanks to Ryan Dew, Jon Jachimowicz, and Yu Ding for all your help, support, and friendship throughout our shared time at Columbia.

DEDICATION

In loving memory of Marisa Castelo.

CHAPTER 1: INTRODUCTION

Consumers see both promise and peril in artificial intelligence (AI). 55% of American consumers say they have used AI, but 42% say they don't trust it (Dujmovic 2017). Similarly, 41% support further development of AI while 22% oppose it (Zhang and Dafoe 2019). The same ambivalence can be seen among business executives, who must decide whether or not to purchase and use AI technologies for their firms: 37% of executives say they have implemented AI in some form, while 20% say they don't trust it (Press 2019). Even AI experts are divided, with 63% predicting that AI would make people better off by 2030 and 37% expecting that it would leave us worse off by that time (Kelleher 2018).

Discussion of both the promises and the perils of AI is often laden with superlatives. Some posit that AI will save hundreds of thousands of lives by removing human error from driving vehicles (Lafrance 2015), diagnosing and treating disease (Perkins 2018), and conducting military operations (Cummings 2017), while also creating trillions of dollars of economic value (Bughin et al. 2018). Others worry that it will destroy the majority of human jobs, create mass unemployment, exacerbate inequality, and even threaten the very survival of the human species should it become more intelligent than we are (Bostrom 2014; Bradshaw 2015). The profound scale of the impact is not often questioned. The desirability of the impact is less clear.

AI's impact will be largely mediated by consumers – both as individuals and as firm employees – deciding whether and when to adopt AI technologies. The goal of this dissertation is therefore to explore how individual consumers perceive and adopt AI. A better conceptual and empirical understanding of this process can provide marketing scholars and practitioners with insight into the dynamics and nuances of AI's impact as well as practical tools to influence the adoption process.

I begin Essay 1 by providing a non-technical description of what AI actually is, breaking down the term into more easily defined and concrete technologies. These include machine learning, knowledge representation and reasoning, natural language processing, computer vision, and robotics. Some combination of these five constituent technologies enable most of the existing applications commonly thought of as AI (also called artificial narrow intelligence or "weak AI"), and together they point to what an artificial general intelligence (or "strong AI") would require in order to become a reality. I then proceed to build a conceptual model of AI adoption, identifying AI's human-likeness, characteristics of the tasks for which AI is used, and consumer heterogeneity as three broad drivers of adoption and providing detailed descriptions and examples of each driver. I also explore the major potential consequences of AI adoption and identify a number of key research questions.

The next two Essays empirically test different parts of the model developed in Essay 1. In Essay 2, I explore how task characteristics influence consumers' willingness to use algorithms to perform tasks normally done by humans, focusing specifically on the perceived objectiveness of the task. I find that consumers mistakenly believe that AI lacks the abilities required to accomplish tasks that seem subjective in nature, leading them to prefer to rely on humans for such tasks even when algorithms perform more effectively. Furthermore, using a combination of lab and field studies, I show that increasing the perceived objectiveness of a task and the perceived affective abilities of the algorithm itself can increase consumers' willingness to rely on algorithms for subjective tasks. The primary contributions of this Essay are (1) the demonstration that willingness to use AI varies substantially depending upon specific features of the task (whereas existing research on this topic has not examined task dependence at all), (2) the finding that increasing the human-likeness of an algorithm can increase the perceived usefulness of the algorithm while also decreasing consumers' comfort with it, illustrating a fundamental tension in AI adoption, and (3) providing insight into how algorithm-based products and services can be advertised in order to increase adoption.

In Essay 3, I focus on consumers' reactions to social robots (physically embodied AI) in consumption contexts. The key finding is that increasing mind perception (the belief that robots can have a human-like mind) improves consumers' reactions to robots with highly human-like physical appearances, both in terms of their perceived usefulness and consumers' comfort with them. Increasing mind perception for robots with a moderately human-like physical appearance, however, decreases comfort while still increasing usefulness. I build on theories of empathy and schema congruity to explain these findings. The primary contributions of this Essay are (1) to demonstrate an interaction between physical and mental human-likeness in shaping consumer reactions to AI, thus providing a new insight into the processes of anthropomorphism and mind perception, (2) creating further evidence for the tension between usefulness and comfort in AI adoption, and (3) suggesting practical tools that the creators and employers of social robots can use to increase the value that they provide both to consumers and firms.

The three Essays together shed light on how consumers perceive and adopt different forms of AI for different purposes. Given the strong ambivalence among consumers regarding these technologies, coupled with the potentially transformative power they hold, the conceptual and empirical progress presented in this Dissertation provide clear value to both marketing scholars and practitioners.

CHAPTER 2: CONCEPTUAL FOUNDATIONS

Consumers and firms both rely on agents to provide services that they value. An agent is anything that perceives input from the environment and then acts upon that environment (Russell and Norvig 2009). Traditionally, these agents have been human beings, whose physical, cognitive, and emotional capabilities allow them to provide a wide range of services. However, a revolution has recently been developing that is dramatically altering this foundation of economic activity. This revolution is artificial intelligence (AI), a collection of technologies that has been endowing machines with increasingly human-like physical, cognitive, and emotional capabilities and thus allowing them to provide services that have traditionally been provided by humans – and often to do so more effectively than humans can. This is already radically changing how consumers and firms obtain services that they value.

This Essay is intended to provide a conceptual structure for understanding what AI is, what it can do, and what are some of the key antecedents and consequences of its adoption and use. The first section will first provide non-technical descriptions of the specific technologies underlying AI, in order to provide a more concrete and detailed understanding of this often-misunderstood concept. The second section will review existing research on the determinants of technology adoption and discuss several limitations that this research faces in explaining the adoption of AI technologies. The third, fourth, and fifth sections will build up a conceptual model of AI adoption, focusing on three key determinants of adoption: characteristics of the technology, of the task for which it is used, and of the consumer. The sixth section explores the consequences of widespread AI adoption, focusing specifically on economic and psychological consequences. The seventh section concludes.

1. SPECIFIC TECHNOLOGIES UNDERLYING AI

There is no widely agreed-upon definition of AI (Stone, Brooks, Brynjolfsson, et al. 2016). It is therefore more useful to think in terms of the more easily definable technologies that are widely agreed to comprise AI. It is first important to distinguish between "artificial narrow intelligence" (ANI, also called "weak AI") and "artificial general intelligence" (AGI, also called "strong AI") (Russell and Norvig 2009). ANI refers to technologies that can perform narrowly defined tasks, such as playing chess or Go, diagnosing diseases, recommending products, driving cars, and so on. AGI refers to a hypothetical technology that would be the equivalent of a human intelligence in terms of its flexibility and capability of performing and learning a vast range of tasks. Many ANIs already exist and are enabled by the specific technologies described below. AGI does not yet exist and experts disagree on whether or not it ever will. In a recent survey of AI researchers, the median estimate was for a 50% chance of achieving an AGI by 2050 and a 90% chance of achieving one by 2075 (Müller and Bostrom 2016).

One definition of AGI is a machine that can pass the Turing Test, proposed by mathematician Alan Turing (Turing 1950). The test is passed if a human interacting with the machine cannot tell whether the responses come from a human or a machine. The original version of the Test involved written responses only, thus requiring that the machine possess natural language processing abilities that allow it to communicate using plain language, plus knowledge representation and reasoning to store and use information. Indeed, natural language processing and knowledge representation and reasoning are two fundamental subfields of modern AI research that will be described below. In order for a machine to pass the so-called "Total Turing Test," however, it is also required to possess human-level perceptual and physical abilities, thus also requiring computer vision and robotics (Harnad 1991). Computer vision and robotics are thus two additional important subfields of AI research that will be described. Finally, perhaps the best-known subfield of AI research is machine learning. Together, a basic understanding of these five concrete and well-defined technologies provides a good introduction to the ANIs already in existence as well as an idea of what a hypothetical AGI would involve.

Machine learning

Machine learning refers to algorithms that can identify patterns in data and then generalize those patterns to make predictions or judgments (Domingos 2012). For example, an algorithm can learn to predict whether a consumer will enjoy different books by being "shown" a dataset consisting of past books that the consumer (or a similar consumer) has read, along with specific features of those books (i.e., genre, length, date of publication, etc.) and how much they were enjoyed (as labeled by humans). The algorithm then observes patterns in the labeled data (often called the training data) and applies those patterns to new books whose enjoyment has not been labeled, allowing it to "predict" whether or not the consumer will enjoy these new books based on the previously learned patterns between book features and rated enjoyment.

Readers may note that this description is essentially similar to linear regression, in which an algorithm "learns" patterns between a set of independent variables (i.e., book features) and a dependent variable (i.e., book enjoyment), which can then be used to predict the dependent variable given a new set of values for the independent variables. Indeed, linear regression is often referred to as a basic machine learning algorithm (Bishop 2006; Murphy 2012). Thinking of machine learning in terms of this basic process of pattern identification and generalization may help understand more complex applications of the technology.

Two key reasons that machine learning algorithms have become widely used in recent years are the increased availability of extremely large datasets and the exponential increase in computing power, which together allow the identification of patterns that could be overlooked by similar algorithms without access to as much data and/or computing power. Pattern identification is facilitated by both dataset length (i.e., a large number of customers in the dataset) and width (i.e., a large number of variables for each customer; Yeomans 2015).

Closely related to machine learning is deep learning, a sub-field of machine learning that uses "neural networks" to identify patterns in data. Neural networks are algorithms whose structure is inspired by the human brain, in which neurons are arranged in layers. Each layer or node of a neural network is itself an algorithm that transforms input data into an output and passes the output onto the next layer of the neural network (LeCun, Bengio, and Hinton 2015). This process is useful for more complex forms of pattern recognition than can be accomplished by traditional machine learning. For example, deep learning can allow computers to recognize handwritten letters by taking an initial data input (pixels), using an initial algorithm (layer) to identify lines and curves, then using a subsequent algorithm (layer) to use the output of the previous layer in order to identify parts of letters, and using a final algorithm (layer) to combine the previous output into complete letters (LeCun et al. 2015).

Like in traditional machine learning, deep learning algorithms are trained on an initial dataset that allows them to learn the accuracy of their predictions. This initial feedback also lets the algorithms adjust the relationships (or "weights") between the layers in order to improve the accuracy of the final prediction. Neural networks can include millions of individual layers and can thus be very costly in terms of both the data and the computing power required to train them (LeCun et al. 2015). These algorithms have applications in fields such as speech recognition (used in Siri and Alexa), computer vision (used in driverless cars and robotics), and other forms of complex pattern recognition (such as detecting tumors), and are particularly useful for modeling large, unstructured datasets (Esteva et al. 2017; Graves, Mohamed, and Hinton 2013; LeCun et al. 2015). Machine learning and deep learning can thus be considered a fundamental AI technology that enables or empowers many of the other constituent AI technologies described below.

Knowledge representation and reasoning

An agent, whether human or artificial, is of course more useful when it knows things and uses that knowledge to do things. Machine learning algorithms represent one way in which machines can acquire knowledge without being explicitly "taught" that knowledge. Machines (such as computers) can also be explicitly taught things, for example by uploading encyclopedias or scientific papers so that the machine stores the uploaded facts in its memory. Regardless of whether the machine learns "on its own" or is explicitly taught, however, it is more useful if the machine can use the resulting knowledge in order to make inferences and reason about the state of the world.

In order to accomplish knowledge representation and reasoning, machines use axioms, sentences, and logic. The axioms and sentences are represented in a knowledge representation language readable by a computer and forming a knowledge base. Artificially intelligent agents can use inference mechanisms in order to infer new knowledge on the basis of its known knowledge. Many of the basic forms of knowledge that humans may take for granted must be understood by an effective artificial agent. For example, basic principles of common sense physics, time, and causality are often required foundations for normal conversations, and effective AI must therefore understand such principles. Similarly, effective AI must also possess ontologies, or formal naming and definitions of categories, properties, and relations between categories and their members (Russell and Norvig 2009). One prominent example of an effort to build a comprehensive ontology understandable by machines is Google's Knowledge Graph, which is a collection and organization of billions of facts used by Google's search engine (Paulheim 2017). More tractable and specific machine-readable ontologies also exist, such as the Gene Ontology project that seeks to teach machines about human genes and their inter-relationships (Padmavathi and Krishnamurthy 2017). Knowledge representation and reasoning is thus a second key technology falling under the AI umbrella.

Natural language processing (NLP)

This third subfield of AI allows artificial agents to recognize, understand, and produce natural language (i.e., language that humans use rather than formal computer programming languages such as Python or C++). NLP involves both written and spoken language. Because natural languages contain ambiguity (i.e., one word with several meanings), and because not all sounds or symbols are meaningful words, NLP relies on the computation of probabilities to understand language. In other words, NLP algorithms determine the probability that a given set of symbols or sounds is a specific word, and the probability that a given word has a particular meaning, given a prior distribution of words or meanings (Russell and Norvig 2009).

Machine learning and deep learning algorithms have been central to the development of NLP in recent years. As described before, such algorithms allow computers to identify specific letters and, in turn, words, by first identifying their constituent features. The same kinds of algorithms can also be used to identify spoken words by first identifying more basic patterns in sound, and underlie technologies such as Amazon's Alexa voice-based assistant. Knowledge representation and reasoning is also an important component of NLP, since it can help an artificial agent to understand basic facts about a natural language, such as the existence of nouns, verbs, and adjectives, and the relationships between those categories of words (Joshi 1991).

Another important part of NLP is natural language generation: writing or speaking language that will be understood by humans (Perera and Nand 2017). Applications of this technology are also found in Alexa-like technologies and in the automatic production of news articles, and even poetry and jokes (Clerwall 2014; Gibbs 2016; Ritchie et al. 2007). NLP therefore allows machines to understand and produce language, both of which are important components of AI.

Computer vision and robotics

As stated earlier, agents perceive inputs from the environment and then acts upon the environment. Artificial agents can "perceive" inputs such as the strokes of a keyboard or the contents of a text file, but also richer inputs such as images of the world itself. Similarly, acting on the environment could be as simple as displaying information on a screen or creating text files, but a richer set of actions is enabled by a physical body that can act on physical objects (Russell and Norvig 2009). Computer vision and robotics thus allow an artificial agent to engage with the world in much more human-like fashion.

Computer vision is a form of perception, in which information about the world is obtained via sensors. In this case, the sensors are often cameras or LIDAR (laser) sensors that use light to create images of the environment. Algorithms then allow the artificial agent to transform the resulting images into a machine-readable format. For example, an image is a collection of pixels, each of which can be coded for color and position using a standardized code in order to represent the image in machine-readable format. Machine learning (and especially deep learning) algorithms are often used to allow the artificial agent to identify the content of images by breaking them down into their constituent components, as described in the machine learning section above. Computer vision can therefore be succinctly described as a combination of sensors that allow a machine to receive inputs of the physical world around it, and algorithms that allow it to interpret the content of those inputs.

Visual inputs are extremely complex. One minute of input from a video camera, for example, can produce 10 gigabytes of data, or about 10,000 times more data than a typical book contains. The useful interpretation of such a huge number of inputs therefore requires the identification of the specific parts of the input that will be most useful for accomplishing the goal in question. Knowledge representation and reasoning can thus become important in the context of computer vision. For example, a driverless car is much more useful if it is able to classify the collection of pixels it perceives as a pedestrian, and so knowledge of pedestrians and their typical behaviors is helpful for knowing how to translate the sensor inputs (i.e., pixels) into actions such as applying the brakes.

Robots are physically embodied agents that can sense and manipulate their environment and perform tasks autonomously, and can be classified into three categories: *manipulators* or stationary robotic arms, such as those used on assembly lines; *mobile robots* that can move around the environment, such as driverless cars or unmanned aerial vehicles (UAVs, or drones); and *humanoid robots*, which mimic the human body and can both move around and manipulate the environment like humans do (Russell and Norvig 2009). Robots use a variety of sensors to receive inputs from the environment, including cameras and microphones, and gyroscopes and accelerometers to measure their own motion. Robots can also use each of the other subfields of AI described above to allow them to learn, interpret inputs, store knowledge, and produce language.

AI can therefore be understood as a collection these five technologies: machine learning, knowledge representation and reasoning, natural language processing, computer vision, and robotics. Each of these technologies enables a wide range of ANI's that perform well-defined tasks but are not easily transferable to other domains. An AGI would require each of these five technologies to operate seamlessly at human levels and to permit flexible learning and transfer across domains and tasks. Having thus gained a better understanding of what exactly AI is, the remainder of the Essay will build a model of how and why consumers adopt and use AI. The next section begins by describing existing attempts to model the process of technology adoption in general.

2. MODELING TECHNOLOGY ADOPTION

There are several existing models of the determinants of technology adoption in the research literature. However, these models were all developed in the context of decades-old technologies whose features and abilities are vastly different from modern AI, suggesting that updated models may be useful for understanding AI adoption. The most well-known academic model of technology adoption is called the Technology Acceptance Model, or TAM (Davis, Bagozzi, and Warshaw 1989). Research on the TAM has primarily focused on employees' use of basic information technologies in the workplace, such as e-mail and word processing software. The TAM posits that two key factors drive individuals' decision to accept and use a new technology: the perceived usefulness of the technology, and the technology's ease of use. An extension of the model (TAM 2) found that perceived usefulness is predicted by both "cognitive instrumental" factors such as result demonstrability, output quality, and job relevance, and by social factors such as image concerns (i.e., whether use of the technology enhances one's social image) and subjective norms (i.e., the perception that others are using the technology) (Venkatesh and Davis 2000).

Several other models have suggested the importance of additional factors in shaping technology adoption. The Theory of Reasoned Action and its extensions to the information technology domain emphasize the importance of attitudes, defined as positive or negative feelings towards the technology, in addition to subjective norms as described in the TAM (Fishbein and Ajzen 1975; Taylor and Todd 1995). The Model of PC Utilization includes several factors also incorporated by the TAM and TAM2, such as usefulness, ease of use, and subjective norms, and additionally emphasizes the role of affect towards the use of the technology (Thompson, Higgins, and Howell 1991). Rogers' Innovation Diffusion Theory includes factors corresponding to usefulness (relative advantage), ease of use (complexity), and subjective norms (observability), and also includes trial-ability (whether the product can be trialed) and compatibility of the technology with existing values and needs (Moore and Benbasat 1991; Rogers 1976). Venkatesh et al. (2003) provides an in-depth review and comparison of these models of technology adoption.

These models have been integrated with TAM and empirically compared in order to yield the Unified Theory of Acceptance and Use of Technology (UTAUT), ultimately providing support for the role of four factors in shaping technology adoption intentions: performance expectancy (corresponding to perceived usefulness), effort expectancy (corresponding to ease of use), social influence, and facilitating conditions (including factors such as having the knowledge and resources required to use the technology; Venkatesh et al. 2003). However, the only one of these factors that had a significant main effect on adoption intentions was performance expectancy (i.e., the perceived usefulness of the technology). The remaining three factors had non-significant main effects but significant interactions with demographic variables such as age and gender.

There are five key limitations that apply to each of these models when considering their application to AI adoption. The first is that all of the research that has tested the models described above has focused on computer technologies that are far simpler than modern AI technologies, such as video conferencing, database and accounting software, personal computers themselves, and so on. This is important because AI is qualitatively different from these and other technologies. Specifically, many applications of AI involve human-level cognitive, emotional, and physical abilities. In other words, AI has *human-likeness*, which is not true of any of the technologies that came before it and which could give rise to an entirely different set of factors influencing its adoption.

The second limitation of existing models arises from the first. Because AI has cognitive, emotional, and physical human-likeness, it can perform a much wider range of tasks than other technologies. Existing research on technology adoption has therefore not explored how different aspects of the task for which a technological product can be used impact consumers' willingness to use that technology for the specific task in question. Exploring whether and why certain tasks are seen as more amenable to the use of AI is therefore an important issue for understanding AI adoption.

The third limitation is that there has been very little research into how consumer heterogeneity impacts technology adoption. Some research has addressed how basic demographic variables such as age and gender moderate the adoption process (Selwyn et al. 2003; Venkatesh and Morris 2000a, 2000b), but there are many other sources of heterogeneity that may be relevant to consumers' attitudes towards technology in general and AI in particular, such as consumer identity, ideology, and overconfidence. The effective marketing of AI technologies will be greatly facilitated by a deeper understanding of how different consumer characteristics impact adoption, both as main effects and as moderators of other factors.

The fourth limitation is that existing models have not sufficiently explored more specific, concrete factors that precede the key drivers of adoption. For example, in the UTAUT, perceived usefulness was identified as the strongest driver of adoption (Venkatesh and Davis 2000), but there was little exploration of what causes a specific use of technology to be seen as useful. This creates both practical challenges for marketers wishing to increase adoption, as well as a conceptual shallowness that hinders a deeper understanding of what really drives technology adoption.

The fifth limitation is adoption is that existing models of technology adoption have ignored the *consequences* of adoption. This is problematic because the widespread adoption of technologies clearly has enormous consequences for the individuals using the technologies, for the firms those individuals make up, and for society in general. A conceptual model that focuses only on the antecedents of AI adoption without addressing the consequences as well is therefore ignoring many of the important research questions on this topic.

The model shown in Figure 1 addresses these five limitations by incorporating human-likeness, task dimensions, and consumer heterogeneity into the conceptual process of AI adoption, by unpacking each of these three broad drivers of adoption into several of their antecedent components, and by addressing some of the most important potential consequences of widespread AI adoption. This unpacking serves to both provide a deeper conceptual understanding of the adoption process as well as to shed light on more practical levers that marketers may use to influence adoption. Each of the three broad drivers and the consequences of adoption will be described in detail in the following four sections.

The model builds on the UTAUT (Venkatesh and Davis 2000), which itself is an integration of several older models including the TAM, Rogers' Innovation Diffusion model, and the Theory of Reasoned Action. Specifically, perceived usefulness is retained as a key driver of adoption from the UTAUT, but is situated as a mechanism through which human-likeness, task dimensions, and consumer heterogeneity impact adoption. Ease of use, social influence, and facilitating conditions did not have main effects on adoption intentions when accounting for perceived usefulness in the empirical estimation of the UTAUT, and so are not included in this model. There are undoubtedly some situations in which these factors do impact adoption. However, because usefulness was by far the strongest determinant of adoption, this model focuses on understanding the antecedents of perceived usefulness.

Furthermore, this model includes comfort as a second mechanism through which human-likeness, task dimensions, and consumer heterogeneity impact AI adoption. Comfort refers to an affective reaction towards the technology that is unrelated to beliefs about the technology's performance or usefulness. The distinction between perceived usefulness and comfort is analogous to similar distinctions in the literature between decision confidence and decision comfort (Parker, Lehmann, and Xie 2016), and between cognitive trust and affective trust (Johnson and Grayson 2005). In each of these distinctions, the basic idea is that judgments and decisions are impacted by both cognitive and affective factors. While the UTAUT found that affective factors such as feelings about the technology were not significant predictors of adoption intentions (Venkatesh et al. 2003), other technology adoption models did emphasize the role of feelings, such as Rogers' Innovation Diffusion Theory (Rogers 1976), the Theory of Reasoned Action (Fishbein and Ajzen 1975), and the Model of PC Utilization (Thompson et al. 1991). Affective factors are expected to play a larger role in the context of AI adoption relative to technologies with no human-likeness. There are several conceptual and empirical reasons why affective reactions (such as comfort) may be more important in the context of AI adoption relative to the adoption of other technologies. These reasons include the existence of strongly negative emotional reactions towards human-like robots (i.e., the Uncanny Valley; Wang, Lilienfeld, and Rochat 2015), the fact that several applications of AI involve emotional abilities or emotional tasks, and the much greater risks that AI adoption poses relative to other technologies (insofar as risk perceptions are largely affect-driven; Loewenstein et al. 2001). Each of these reasons and others will be discussed in greater detail in subsequent sections.

The proposed model therefore identifies consumers' comfort with AI and their perceived usefulness of the technology as the two proximate determinants of adoption, and identifies AI human-likeness, dimensions of the task for which the AI is used, and consumer heterogeneity as three broad classes of variables that in turn shape both comfort and perceived usefulness. Insofar as the technology adoption process consists of the technology, the consumer adopting it, and the use of the technology for a specific task, these three classes of variables represent each of these three components of the process. AI's human-likeness represents key characteristics of the technology, and consists of physical appearance, cognitive abilities, and emotional abilities. Consumer heterogeneity represents key characteristics of the consumer adopting the technology; some of the most relevant sources of this heterogeneity to AI adoption are consumers' overconfidence, desire for control, trust in feelings, the extent to which they identify with the task for which AI is being used, their conservatism, age, and gender. Task dimensions represent key characteristics of what the technology is used for; some of the most relevant task dimensions to AI adoption are the economic context of the task, including the price of using AI relative to using humans to perform the task, how risky the task seems, the affect or emotion involved in the task, and how enjoyable the task is for consumers. Each of these three classes of variables will be detailed in the following sections, beginning with the concept of AI human-likeness.

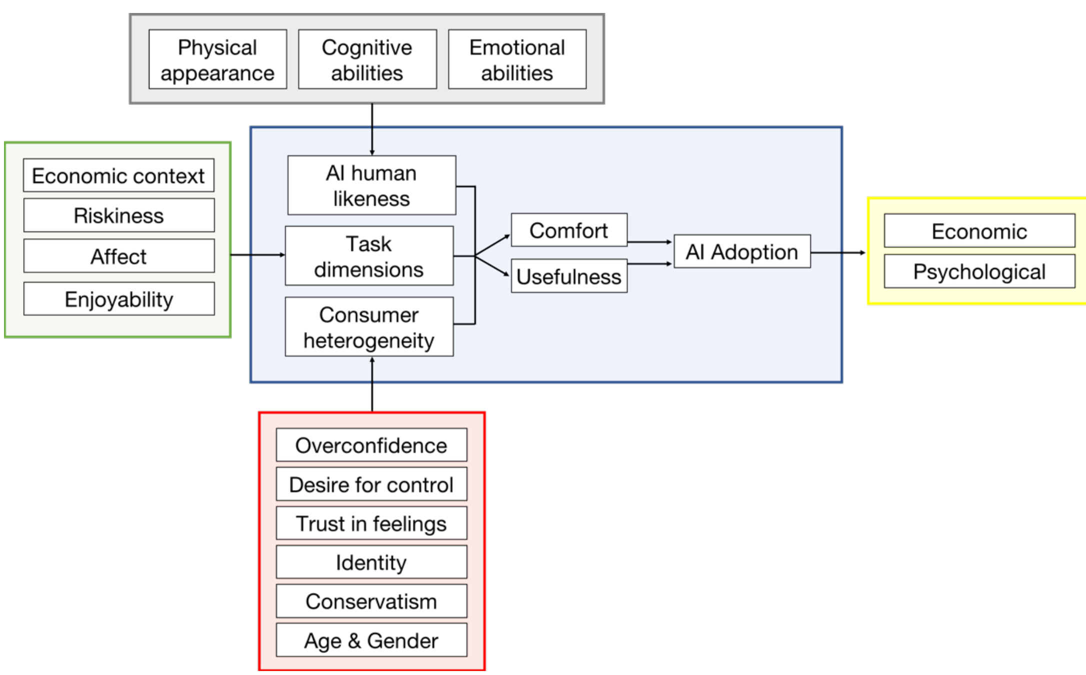


Figure 1. Conceptual model of AI adoption.

3. AI HUMAN-LIKENESS

As AI continues to progress, it acquires increasingly human-like abilities. The specific technologies described in Section 1 allow AI to learn a vast array of information and skills, to store and reason with knowledge, to understand and produce hundreds of human languages, and to see and interact physically with the environment. AI's *human-likeness* is thus clearly increasing along many dimensions. However, progress is faster along certain dimensions, and consumers and managers are likely to perceive AI-enabled products differently depending upon what dimensions of human-likeness they display. Increasing human-likeness along certain dimensions may increase perceived usefulness while decreasing comfort, for example, while increasing human-likeness along another dimension may simply increase perceived usefulness without affecting (or even increasing) comfort. Furthermore, different dimensions of human-likeness also enable AI to perform different kinds of tasks, which in turn may also affect consumers' perceptions of the technology and its applications. An understanding of the dimensional structure of human-likeness is therefore useful for understanding AI adoption.

Several literatures in psychology have explored dimensions of what might be called human-likeness, albeit with different terminology depending on the particular literature. These literatures each converge on two dimensions, roughly corresponding to emotional and cognitive abilities. For example, research on dehumanization has shown that people perceive two categories of human abilities. First are "human uniqueness abilities," which distinguish humans from other animals but can be shared with machines. These tend to be cognitive in nature (such as logic and rationality). Second are "human nature abilities," which may be shared with other animals but not with machines. These tend to be affective or emotional in nature (such as warmth and intuition; Haslam 2006; Loughnan and Haslam 2007). Importantly, research has shown that machines such as robots are seen as lacking human nature abilities (which are emotional) but not human uniqueness abilities (which are cognitive; Haslam et al. 2008).

Research on mind perception has focused on two similar dimensions: agency, or the ability to engage in intentional planning and action, and experience, or the ability to subjectively experience emotions and sensations. Mirroring research on humanness, machines (such as robots) are seen as having some degree of agency but no experience (Gray, Gray, and Wegner 2007), and endowing robots with experience creates more negative reactions than endowing robots with agency (Gray and Wegner 2012).

Similarly, research on person perception has identified the two dimensions of competence and warmth as fundamental to how we perceive other people (Fiske, Cuddy, and Glick 2007). These two dimensions also correspond to more cognitive and emotional abilities respectively. These three streams of research thus demonstrate that consumers perceive human abilities as either cognitive or emotional and are willing to grant machines more cognitive than emotional abilities. The first two dimensions of human-likeness can therefore be called cognitive and affective human-likeness, which together make up a human-like mind.

Cognitive human-likeness

AI is providing machines with the ability to perform both cognitive and emotional tasks. The most familiar applications of AI today tend to be more cognitive in nature: using machine learning to recognize patterns in large datasets, making predictions about future outcomes on the basis of prior observations, analyzing and summarizing data, and so on. However, there are many other "cognitive" abilities that remain out of reach for AI. For example, explaining the reasons underlying a given decision or prediction remains a major challenge for machine learning algorithms whose precise formulation is often opaque even to the creator of the algorithm (Dosilovic, Brcic, and Hlupic 2018; Park et al. 2017). Many of these algorithms use so many factors as inputs, and so many layers separating the inputs from the output in the case of deep learning algorithms, that providing a clear explanation of how the output was produced is not feasible. There has therefore been a recent surge of interest in creating "explainable AI," meaning algorithms that can explain how they arrived at a given decision, such as diagnosing a disease or approving a loan (Dosilovic et al. 2018).

This effort is also related to the important issue of algorithmic bias: many algorithms are trained on datasets that reflect human bias, which then becomes formally entrenched in the algorithm itself (O'Neil 2016). For example, an algorithm used to evaluate job candidates may be trained on a dataset in which most "successful" candidates are male and therefore penalizes female applications, or a facial recognition algorithm may be trained on a dataset in which most of the faces are Caucasian and therefore is less effective at recognizing non-Caucasian faces (Buolamwini and Gebru 2018). Building algorithms that can transparently explain their decisions is therefore a crucial step towards fair, unbiased algorithmic decision making – and therefore also towards unbiased cognitive human-likeness.

A second major hurdle in achieving cognitive human-likeness is the kind of cognitive flexibility that allows humans to transfer knowledge and experience from one domain to another. This reflects the distinction introduced earlier between artificial narrow intelligence (ANI) and artificial general intelligence (AGI). All existing forms of AI are narrow, in the sense that they are trained in one specific domain such as playing a specific game, diagnosing a disease, or driving a car, but are unable to function in other domains (Russell and Norvig 2009). There is progress being made in allowing algorithms to transfer knowledge and abilities from one domain to another, although the level of flexibility is still extremely limited relative to humans (Taylor and Stone 2007; Weiss, Khoshgoftaar, and Wang 2016).

Despite these limitations, however, AI's human-likeness is clearly higher along the cognitive (vs. affective) dimension: the growing list of tasks that AI can perform at human-levels consists mostly of non-affective prediction and classification tasks (Grove et al. 2000; Castelo, Lehmann, and Bos 2019), and AI has more "human uniqueness" and "agency" abilities (the cognitive abilities identified as central to humanness in the dehumanization and mind perception literatures) than it has "human nature" and "experience abilities (the affective abilities identified as central to humanness in those literatures; see Table 1).

High cognitive human-likeness is compatible with the stereotypical association of computers with cognitive abilities (Gray and Wegner 2012; Loughnan and Haslam 2007), and increasing cognitive human-likeness in AI should therefore be perceived relatively positively, as category-congruent or stereotype-congruent examples usually are (Loken 2006; Meyers-Levy and Tybout 2002). Specifically, increasing cognitive human-likeness will likely increase the perceived usefulness of AI without producing negative affective reactions such as discomfort.

However, if cognitive human-likeness eventually approaches perfectly human levels in every sense – i.e., not only in certain narrowly defined cognitive tasks but with the full explain-ability and flexibility of human cognition – this level of cognitive human-likeness may start to create discomfort, potentially stemming from perceived threats to human jobs, human distinctiveness or even to human safety. Indeed, the threat that AI poses to human jobs and even to human safety is a frequent narrative in the news media (Bradshaw 2015; Cellan-Jones 2014; Lohr 2018), and there is some evidence that robots are seen as threatening human distinctiveness from machines, which contributes to negative evaluations of robots (Ferrari, Paladino, and Jetten 2016). Thus, the effects of cognitive human-likeness on comfort may be non-linear, with initial increases preserving comfort with the technology, but further increases creating perceived threats and discomfort even as the technology becomes more useful. The non-linear effect of *physical* human-likeness on reactions to robots in particular has been explored in the literature on the Uncanny Valley hypothesis, which will be described in the section below on physical human-likeness. This suggests that increasing AI's cognitive human-likeness will have positive linear effects on perceived usefulness, but non-linear effects (initially positive, then negative) on consumers' comfort.

Cognitive Abilities		Affective Abilities	
Human Uniqueness	Agency	Human Nature	Experience
Rationality	**Memory**	**Emotional responsivity**	Hunger
Cognitive sophistication	**Emotion recognition**	Interpersonal warmth	Fear
Morality	**Planning**	Cognitive openness	Pain
Civility	**Communication**	Individuality	Rage
Refinement	*Thought*	Depth	Desire
Maturity	*Self-control*		Personality
	Morality		Consciousness
			Pride
			Embarrassment
			Joy

Table 1. AI's cognitive and affective abilities. Note: AI has the abilities in bold and arguably has the abilities in italics.

Affective human-likeness

Turning to the affective dimension of human-likeness, there are a growing number of tasks typically associated with emotion that AI can now perform. This includes most prominently the ability to accurately recognize and distinguish between different emotions being expressed in human faces, voices, and writing (Liu 2010; McDuff et al. 2013; Picard 2011). Note that despite emotion recognition being categorized as an "agency" ability in the mind perception literature (Gray, Gray, and Wegner 2007), the fact that it explicitly involves emotions should result in it being classified by lay consumers as a relatively more emotional (vs. cognitive) ability. AI's affective abilities also include emotional responsivity, or the ability to alter its' behavior depending on how the humans with whom it interacts are feeling, which is classified as an affective ability in the dehumanization literature.

Affective human-likeness in AI also includes the abilities to create paintings that sell for hundreds of thousands of dollars (Quackenbush 2018) and write compelling poetry and music (Deahl 2018; Gibbs 2016b). The processes that AI uses to perform these tasks may appear very different from the processes that humans use (i.e., deterministic computer code rather than intuitions or gut feelings), although there are deep philosophical questions about whether such intuitions and gut feelings are also the result of deterministic physical processes in the brain, analogous to the physical processes on a silicon computer chip (Greene and Cohen 2012; Nichols 2008). Regardless of the mechanisms with which humans and computers perform these tasks, AI can indeed perform some tasks normally thought of as emotional or affective, which should increase the degree to which AI seems human-like along the affective dimension of human-likeness.

Overall, however, AI is clearly very low in affective human-likeness (see Table 1). High affective human-likeness is not compatible with the stereotypical associations that people have with computers (Gray and Wegner 2012; Loughnan and Haslam 2007) and is therefore expected to elicit negative affective reactions. Indeed, this dimension of human-likeness is often considered to be the key distinction between humans and machines, and denying humans affective abilities is called mechanistic dehumanization – likening the human to an unfeeling machine (Haslam and Loughnan 2014). Endowing AI with affective human-likeness is therefore expected to be particularly threatening to human distinctiveness and should thus create more discomfort than endowing AI with cognitive human-likeness, despite also making AI seem more useful. This suggests that increasing AI's affective human-likeness may have positive linear effects on perceived usefulness, but linear negative effects on consumers' comfort.

For both cognitive and affective human-likeness, discomfort stemming from perceived threats to humans may be at least partially offset by increases in the perceived usefulness of AI with high human-likeness. However, this tradeoff between comfort and usefulness is expected to differ depending on the dimension of human-likeness, with usefulness playing a relatively smaller role in shaping adoption of AI with affective (vs. cognitive) human-likeness due to greater discomfort with affective AI.

AI is therefore acquiring an increasingly human-like mind, although progress is faster along the cognitive dimension. Further increases along both dimensions may increase the perceived usefulness of the technology and thereby increase adoption, although may also produce discomfort and thereby decrease adoption. The effects of human-likeness on comfort may be non-linear and may be particularly strong in the context of affective human-likeness. Finally, the effects of cognitive and affective human-likeness on both usefulness and comfort are likely to depend on the specific task for which the AI is being used: this issue is explored in greater detail in the following section on Task Dimensions.

Physical human-likeness

Humans are not just minds – they also have bodies. In addition to human-likeness varying along the two primary dimensions of mind, human-likeness must therefore also include a physical dimension. This dimension becomes particularly relevant when AI is physically embodied – i.e., in the case of robotics. Robots vary significantly in terms of their physical human-likeness, or how human-like they look. Industrial robots used in factories tend to have extremely low human-likeness, perhaps slightly resembling a human arm. Most consumer-facing robots at least allude to human features by using a large round screen for a face but lack any recognizably human body, such as Jibo (sold for use in the home and recently featured on the cover of Time magazine as one of the best inventions of 2017). Other robots such as Pepper (used in retail stores and restaurants) are more human-like in their appearance, having eyes and a mouth as well as extremities resembling human legs, arms, and hands. Finally, some robots such as Sophia (who appears as a speaker at conferences and as a guest on talk shows) and Erica (employed as a news anchor in Japan) are designed to be exact replicas of humans and are increasingly difficult to distinguish from real humans.

Research using a database of 200 robots asked participants to indicate the presence or absence of several human features on each robot, and used principal components analysis to show that physical human-likeness can be summarized by three primary factors: surface appearance, which includes eyelashes, hair, skin, genderedness, eyebrows, and apparel; body appearance, which includes hands, arms, torso, fingers, and legs; and facial appearance, which includes eyes, mouth, head, and face (Phillips et al. 2018). In order, the strongest predictors of overall human-likeness were surface appearance, body appearance, and facial appearance.

High physical human-likeness in robots often elicit strong affective reactions, famously described by the Uncanny Valley hypothesis. Nearly 50 years ago, Masahiro Mori, a Japanese roboticist, wrote an influential paper speculating that making robots look more human-like is beneficial only up to a point, after which they become *too* human-like and elicit strongly negative reactions (Mori 1970). These reactions are often described as "creeped out," "unnerved," or "eerie," which are clearly affective reactions having little to do with the usefulness of such robots. Instead, something about a machine becoming too human-like in terms of its' physical appearance seems to be deeply unsettling (Wang et al. 2015). Research has suggested that one source of this discomfort is the notion that human-like robots challenge the belief that humans are distinct from machines, and that this lack of distinctiveness is itself upsetting (Ferrari et al. 2016). Thus, high human-likeness can create strong affective reactions not captured by the existing models of technology adoption that focus predominantly on the usefulness of the technologies. Furthermore, as mentioned above, the non-linear effect of physical human-likeness documented in research on the uncanny valley is expected to also be observed in the context of cognitive and affective human-likeness, although this remains an open question for future research.

One open question in this context is how consumers will react to robots with *perfect* physical human-likeness. On one hand, many of the recognized sources of discomfort with highly (but imperfectly) human-like robots would be eliminated, such as the aesthetic imperfections of such robots that increase mortality or disease salience (Ho, MacDorman, and

Pramono 2008; MacDorman et al. 2009). On the other hand, other recognized sources of discomfort with human-like robots would remain, and perhaps be amplified, such as category uncertainty (is this a human or a robot?) and perceived threats to human distinctiveness (Ferrari et al. 2016). Because some sources of discomfort would thus be eliminated while others would remain, perfectly human-like robots should be expected to fall in between imperfect robots and humans in terms of the affective reactions they elicit.

The effect of physical human-likeness on perceived usefulness remains an open question. Robots with very human-like bodies (i.e., functioning limbs that allow them to walk, carry objects, and so on) are of course objectively more useful than robots with less human-like bodies, at least for performing certain tasks. The practical purpose of having a very human-like face is less clear, although it could make robots more useful for tasks involving a social dimension. These questions will also be addressed in more depth in the Task Dimensions section.

There are also likely to be interactions between the dimensions of human-likeness. One of the explanations for the Uncanny Valley phenomenon is known as the perceptual mismatch hypothesis, and states that discomfort with human-like robots is caused in part by a mismatch between different part of the robot’s human-likeness, such as a human-like face with a robotic-sounding voice, or a human-like body with robotic movements (Kätsyri et al. 2015). A related explanation, known as the *violation of expectations hypothesis*, which was the explanation proposed by Mori in the first Uncanny Valley paper, argues that human-like robots create an initial expectation of a human but then fails to meet those expectations, which creates discomfort (Mori 1970). Both of these explanations suggest that congruence between the dimensions of human-likeness can alleviate discomfort, such that potentially negative effects of increasing a given dimension of human-likeness in isolation may be reduced when increasing the other dimensions simultaneously. Thus, a robot with a human-like physical appearance may be expected to also think like a human does, while a robot that thinks like a human may also be expected to understand and express emotions like humans do. Thus, acknowledging that increasing human-likeness may have non-linear and negative effects on comfort, *ceteris paribus*, congruence among the dimensions of human-likeness should increase comfort.

Finally, there is an apparent tension between usefulness and comfort in the context of all three dimensions of human-likeness. As human-likeness increases, AI becomes objectively more useful, at least for certain tasks, thus increasing the likelihood of adoption. At the same time, however, such increases can also produce discomfort among consumers, thus decreasing the likelihood of adoption. Understanding how consumers manage this tradeoff and how marketers can optimize it is therefore a key question for research on AI adoption.

4. TASK DIMENSIONS

The fact that AI has some degree of cognitive, affective, and physical human-likeness enables the technology to perform a wide range of different tasks. In contrast, most of the technologies studied in existing models of technology adoption are only capable of performing one narrow set of tasks, such as word processing or video conferencing. The range of tasks for which AI can be used vary along many dimensions that are likely relevant to consumers' willingness to use AI for a given task. Empirical research on this question is virtually non-existent, although one paper has explored how task objectivity impact trust in and reliance on algorithms (Castelo, Bos, and Lehmann 2019), and some of the other relevant dimensions can be identified conceptually. The purpose of this section is not to identify all dimensions along which tasks can vary – such a list would be infinitely long. Instead, the goal to identify a few such dimensions clearly related to AI adoption in order to provide a starting point from which future research can further explore this broad driver of adoption.

Task affect

First, tasks themselves can be construed as being more or less “affective” or “cognitive.” For example, some tasks are seen as more objective in nature, meaning that they are rule- and logic-based, while others are seen as more subjective, meaning that they are intuition- or feeling-based (Inbar, Cone, and Gilovich 2010). Research has found that consumers are less trusting and willing to use algorithms for tasks that seem more subjective in nature, because algorithms are thought to lack the affective abilities required for subjective tasks (Castelo, Bos, and Lehmann 2019). In other words, if the task for which AI is being used is believed to involve affective abilities or qualities, then consumers may be less likely to use AI for that task, relative to tasks that do not involve such abilities or qualities, because AI is believed to be less useful for such tasks. In that paper, usefulness was found to be a stronger predictor of actual reliance on algorithms compared to comfort, although comfort had significant effects without controlling for usefulness. This suggests that any discomfort consumers’ feel with algorithms being used for certain tasks can be potentially offset if the algorithm is very useful.

This logic can be extended to related task dimensions beyond task objectiveness. For example, the degree to which a task is seen as hedonic vs. utilitarian is also related to how much affect is involved in the task. Hedonic tasks involve affective, sensory, and aesthetic factors relative to utilitarian tasks which involve more cognitive, instrumental, and functional factors (Dhar and Wertenbroch 2000; Holbrook and Hirschman 1982; Shiv and Fedorikhin 1999).

Holding all other factors equal, AI should thus be seen as less useful for tasks that involve a hedonic component, relative to tasks that are purely utilitarian in nature.

Also related to task affect is the degree to which a task involves a social component. Emotions and social interaction are fundamentally intertwined. Emotions are often the *result* of social interaction; emotions have social *consequences* for other people; emotions *facilitate* social interaction (Keltner and Haidt 1999; Parkinson 1996). AI, which is accurately seen as relatively lacking in affective human-likeness, should therefore be seen as less effective for tasks involving social interaction, and consumers may be less comfortable with the use of AI for such tasks.

These propositions clearly suggest an interaction between task and human-likeness. Consumers have strong lay theories about what kinds of tasks require what kinds of mental abilities (Inbar et al. 2010). This leads to AI being seen as less useful for tasks requiring the abilities that AI is perceived to lack. Increasing the dimension of AI human-likeness that corresponds to the abilities required for a certain task should therefore increase AI's perceived usefulness for the task, increasing the likelihood of adoption.

This logic is less straightforward when it comes to physical human-likeness. However, because facial expressions play an important role in facilitating social interactions (Frith 2009), AI's physical human-likeness may also interact with the degree of social interaction involved in the task for which it is used, increasing AI's physical (specifically facial) human-likeness may increase consumers' comfort with and perceived usefulness of AI for tasks involving social interaction, more so than for tasks not involving social interaction.

These propositions help to illustrate the value of the proposed adoption model by highlighting not only the main effects of human-likeness and task dimensions, but also the ways in which they are likely to interact in shaping AI adoption via the mechanisms of usefulness and comfort.

Task riskiness

Beyond the affective nature of the task, other factors are also relevant in shaping consumers' willingness to use AI to perform a task. One such factor is the perceived riskiness of the task. Risk perceptions are a major determinant of attitudes towards new technologies (Slovic, Fischhoff, and Lichtenstein 1982), although they have been neglected by the models of technology adoption reviewed earlier, likely because the technologies addressed in such models are relatively risk-free. Risk perceptions are largely determined by two factors: the consequentialness of the task and the probability of the outcome. In other words, something's perceived risk is a function of the *importance of its potential consequences* (i.e., its consequentialness), multiplied by the *likelihood of those consequences occurring* (Bettman 1973; Jacoby and Kaplan 1972).

The many tasks that AI can perform vary significantly in terms of their consequentialness. For example, using AI to recommend a movie on Netflix is inconsequential relative to using AI to drive a car or diagnose a disease. Using AI for more consequential tasks should therefore be seen as riskier, which should in turn decrease consumers' willingness to use AI for such tasks. Indeed, research has found support for this hypothesis, especially among more conservative consumers for whom risks are more salient (Castelo, Bos, and Lehmann 2019; Castelo and Ward 2019).

In addition to consequentialness of the outcome, the second major determinant of risk perception is the probability of a given outcome occurring. Unlike consequentialness, this factor is not inherent in the task itself, but is shaped by considerations such as belief in the technology's effectiveness and reliability, and trust in the designer and operator of the technology (Lee and See 2004). For example, a technology with a long track-record of success should increase consumers' belief that the technology can perform well and reliably, thus decreasing the probability that negative outcomes will occur and decreasing perceived riskiness. Similarly, if the technology is designed and operated by a company or government that a consumer trusts, the perceived riskiness of the technology should also decrease.

Risk perceptions are also shaped by many well-known heuristics and biases, such as the availability heuristic and the affect heuristic (Folkes 1988; Slovic, Fischhoff, and Lichtenstein 1981; Slovic and Peters 2006). Thus, both perceived consequentialness and probability of failure may be distorted by salient events such as the first fatal accident caused by a driverless car, ultimately increasing the perceived risk of such technologies (Shariff, Bonnefon, and Rahwan 2017). Popular media narratives also likely fuel the perceived riskiness of AI, including the notion that AI will take over human jobs (Lohr 2018) and even pose an existential risk to humanity (Bradshaw 2015). Risk-as-feelings, or the notion that perceived risk can be largely driven by one's feelings (Loewenstein et al. 2001), is another key reason why incorporating affective reactions into a model of AI adoption is important, especially relative to existing models of technology adoption that address technologies in which risk is less salient.

The riskiness of a given application of AI should primarily impact consumers' comfort with the technology rather than its' perceived usefulness – although motivated reasoning (Kunda 1990) may lead consumers who are uncomfortable with a risky application to also question whether the technology itself is useful for the application in question. This notion also points to a potential relationship between comfort and usefulness, such that consumers who feel less comfortable with a particular use of AI may be less likely to believe that AI is useful for that task.

Task enjoyability

The enjoyability of a task may also impact consumers' willingness to use AI to perform the task. This idea will be explored in more detail in the following section on consumer heterogeneity, since task enjoyability is likely to vary substantially depending on the consumer in question and his or her preferences and identity. The basic intuition, however, is that consumers may be less willing to use AI to perform tasks that they find enjoyable, even if AI can perform the task more efficiently or accurately than they can themselves. For example, recent research has found that consumers are reluctant to use AI to help them recommend jokes to other people, despite knowing that the AI can do so more effectively than they can (Yeomans et al. 2019). One possible source of this reluctance is that reading and recommending jokes is an enjoyable task from which consumers obtain some utility. To the extent that consumers receive utility from providing recommendations to others and thereby expressing their likes and dislikes, they may be less willing to use AI to provide recommendations for a wide range of products and services. Similarly, consumers who find driving very enjoyable may be especially unlikely to purchase a driverless car (Leung, Paolacci, and Puntoni 2018).

The enjoyability of a task may itself be determined by the extent to which is provides happiness and/or meaningfulness (Baumeister et al. 2013). Thus, even if a task is not strictly "fun" in the sense of providing immediate happiness or pleasure, it may nevertheless provide consumers with a sense of meaning and therefore makes the task more enjoyable in a different sense.

Consumers may be *more* willing to use AI to perform tasks that they find unenjoyable. The widespread use of robotic vacuum cleaners such as the Roomba and of algorithm-based tax preparation software such as TurboTax are two examples of rudimentary AI products commonly used to automate unenjoyable tasks.

Economic context

The economic context of a given task is also highly relevant to the use of AI for that task. For both individual consumers as well as for firms, the economic context includes the relative cost of using a human vs. using AI for a given task. While consumers and firms can and do employ other humans to perform many tasks, the declining financial cost of using AI instead makes it increasingly feasible to outsource a growing number of tasks to machines. As with most other products and services, declining prices for AI technologies will lead to increasing use.

The economic reality of adopting AI technologies is likely to be more complex for firms than for individual consumers. Consumers can purchase ready-made, AI-enabled products or services such as recommendation software or autonomous vehicles, without needing any technical expertise to use their purchase. On the other hand, firms interested in incorporating AI into their business practice or creating AI-enabled products must also have employees with the required technical expertise. Such employees are in high demand: only 22,000 people worldwide have PhD-level training in AI (Kahn 2018) and the average employee in a top AI lab commands a salary of $345,000 (Metz 2017). While many AI applications can be implemented with much less than PhD-level training, the intense competition for human talent nevertheless complicates the economic analysis for firms wanting to use AI. Firms located in countries that invest heavily in training AI researchers, such as the US, China, the UK, and Canada, will have an advantage in this competition (Manyika and Bughin 2018).

In addition to the cost of hiring the human programmers required to develop and maintain AI technologies, another economic consideration for firms is the relative cost of automating different kinds of jobs and tasks. Repetitive physical tasks such as assembly line work, food production, cleaning, and so on are technically easier (and therefore less costly) to automate relative to jobs that involve social and emotional skills, and so firms should be expected to automate the first kind of tasks sooner (Manyika, Chui, et al. 2017). This reflects the fact that AI's affective human-likeness remains lower than it's cognitive and physical human-likeness. However, the broader economic context in which a firm is situated is again expected to play a role: countries in which human labor is inexpensive will be less likely to adopt AI to replace human jobs (including even repetitive physical jobs), relative to countries in which human labor is more expensive (Manyika, Lund, et al. 2017).

The task dimensions highlighted in this section (the affect involved in the task, riskiness, enjoyability, and economic context) were described mainly as main effects – in other words, these dimensions are expected to impact AI adoption holding the other dimensions constant. However, there are likely to be interactions among these dimensions. Task subjectivity may matter less when consequentialness is low, for example. It is therefore important for future research to explore how these and other dimensions interact in nuanced ways in addition to identifying main effects. The relative impact and potential interaction between usefulness and comfort as mechanisms of these effects is also an important question in the context of task dimensions, as it is in the context of human-likeness.

5. CONSUMER HETEROGENEITY

The previous two sections focused on how characteristics of AI itself (i.e., its' human-likeness) and of the tasks for which AI is used are likely to impact the adoption process. A third set of factors relevant to this process are characteristics of the consumer. This section is also not intended to identify *all* relevant sources of consumer heterogeneity; such a list is potentially infinite. Instead, the purpose is to identify some of the most important relevant sources of heterogeneity based on existing research and conceptual analysis.

Demographics

The most basic source of consumer heterogeneity relevant to AI adoption is demographic heterogeneity. Age and gender in particular have been investigated in the context of existing technology adoption models, although mostly as moderators of other factors rather than as main effects. For example, research has found that men's technology adoption is impacted more by usefulness, while women's is impact more by ease of use (Venkatesh and Morris 2000), and that older adults are more influenced by subjective norms (Venkatesh and Morris 2000b). Beyond the literature on technology adoption models, main effects of age have also been observed such that older adults are less likely to adopt and use technological products (Meyer 2011; Selwyn et al. 2003).

Research on consumers' trust in AI in particular has found a consistent and large main effect of gender, such that women are less trusting of AI for performing consequential tasks such as driving cars and diagnosing diseases but not for performing less consequential tasks such as recommending a movie (Castelo and Ward 2019). As described in Section 4, consequentialness is a major determinant of perceived risk (Bettman 1973; Jacoby and Kaplan 1972), and females are known to perceive more risk in general (Gustafson 1998) and to take fewer risks than men (Byrnes, Miller, and Schafer 1999). This may explain why gender impacts perceptions of certain technologies (i.e., consequential applications of AI) but not others.

Identity

Beyond demographic variables, research has also found that consumers' identification with a given task impacts their willingness to automate that task, which has implications for AI adoption. For example, if driving cars is an important part of a consumers' identity, then that consumer may be less willing to purchase a car with automatic transmission (Leung et al. 2018). This is related to the task enjoyability point made in Section 4. When a task is an important part of a consumer's identity, they likely enjoy performing that task, and are thus less likely to want to automate that task using AI. Thus, consumers who identify as passionate drivers may also be less likely to purchase a driverless car.

Identity may also be a relevant factor in the context of workplace adoption of AI technologies. AI can already effectively diagnose diseases and write newspaper articles, but doctors and journalists may be reluctant to use these technologies if doing so is perceived as threatening their professional identities, or even their livelihood itself. Indeed, people's job often makes up an important part of their identity (Rosso, Dekas, and Wrzesniewski 2010), implying a strong motivation to maintain that identity. This motivation can be both intrinsic (i.e., to view oneself as competent and useful) and extrinsic (i.e., to be viewed as competent and useful to others). Technologies that threaten one's sense of competence and usefulness at work may therefore be discounted.

Identity-related concerns may also be relevant to a broader sense in which AI can be threatening. Consumers tend to value their membership in distinct groups because this helps to create a sense of identity, and when outgroup members threaten the distinctiveness of their group, the outgroup can be evaluated negatively (Tajfel and Turner 1986). A prominent example of this phenomenon is humanity (in which all humans are members) becoming less distinct as AI becomes more human-like. To the extent that being human is an important part of one's identity, increasingly human-like AI may be seen as threatening to one's identity, and this perceived threat may in turn decrease the likelihood of adopting AI.

Trust in feelings

Consumers' trust in feelings as an input to decision making is another likely relevant factor shaping AI adoption. Consumers differ in the extent to which they believe that their feelings are "trustworthy" and lead to good judgments and decisions, and higher trust in feelings leads to greater reliance on feelings as a judgment criterion (Avnet, Pham, and Stephen 2012). Consumers who trust and rely more on their feelings when making decisions may be more strongly influenced by affective reactions to AI such as discomfort with the idea of relying on a human-like machine, and less strongly influenced by more utilitarian factors such as perceived usefulness or economic incentives.

These ideas can be applied to individual consumers or groups of consumers (i.e., cultural groups) who vary in their tendency to trust their feelings when deciding. Interestingly, they may also be relevant to a comparison between B2B vs. B2C applications of AI. Specifically, businesses may be more likely to make AI adoption decisions on a purely "cognitive" basis, with fewer opportunities for feelings to impact such decisions compared to individual consumers. Indeed, if a given application of AI would increase a company's profitability, that company may have a fiduciary duty to adopt it, and such adoption decisions would therefore not be influenced by affective processes. In contrast, individual consumers tend to have much less formalized or institutionalized decision-making processes than firms do, leaving open far more opportunities for affect to influence the decision-making process. This may be one reason why businesses are likely to adopt AI technologies more quickly than individual consumers.

Overconfidence and desire for control

Two final and closely related sources of consumer heterogeneity are overconfidence and desire for control. The latter concept refers to the desire to personally exert control over one's environment (Leotti, Iyengar, and Ochsner 2010) and has been linked to a decreased interest in adopting new products because such products often entail changing routine behaviors and do not fit in to existing categories or schemas (Faraji-Rad, Melumad, and Johar 2017). Similarly, providing consumers with control over an algorithm's output (i.e., allowing them to modify its output) increases willingness to rely on the algorithm (Dietvorst, Simmons, and Massey 2016), providing further support for the role of desire for control. Desire for control may itself be increased among consumers who identify with or enjoy the task in question and who are more risk-averse.

One additional important reason why consumers may wish to control outcomes that could instead be automated by AI is overconfidence. In other words, consumers may believe that they can perform the task better than they really can, or, importantly, better than an algorithm. People tend to be overconfident both in the sense that they overestimate their actual performance and mistakenly believe they perform better than others (Moore and Healy 2008). Overconfidence has been linked to suboptimal investment strategies among CEOs (Malmendier and Tate 2005) and to the creation of speculative asset bubbles (Scheinkman and Xiong 2003). Furthermore, expert national security analysts (but not non-experts) discount accurate advice from algorithms, suggesting that experts in particular may be prone to overconfidence in their abilities relative to algorithms' abilities (Logg, Minson, and Moore 2019). Thus, a desire to maintain control over tasks and decisions, stemming partly from overconfidence, may also contribute to reluctance to adopt AI.

However, expertise may have divergent effects on AI adoption depending on the type of expertise in question. Technical expertise with AI itself is likely to increase adoption likelihood, because AI experts are able to better understand how the technology works, thus potentially increasing its' perceived usefulness (Venkatesh et al. 2003; Yeomans et al. 2019). In contrast, having expertise in the *task* for which AI is being used, but not with AI itself, may decrease willingness to adopt AI by increasing overconfidence with one's own performance in the task (Logg, Minson, and Moore 2019).

As with the task dimensions described in the previous section, this list of sources of consumer heterogeneity is also not meant to be exhaustive, but to identify some of the most important sources and thus to provide a starting point for thinking about and researching this broad source of factors that impact AI adoption.

6. CONSEQUENCES OF AI ADOPTION

Understanding the antecedents of adoption without a good understanding of the consequences is risky, especially if marketers and policymakers manipulate the antecedents to affect adoption without awareness of the impact that such changes could bring. Widespread adoption of AI technologies can be expected to have profound economic, political, and psychological consequences, although none of these consequences are yet well understood. This section will focus on the potential negative consequences of AI adoption, since the positive consequences that AI offers are relatively straightforward (i.e., performing tasks more effectively and for a lower cost than humans can).

Economic consequences

Perhaps the largest potential consequences of widespread AI adoption will be economic in nature. While AI adoption will likely benefit firms by increasing their productivity (Bughin et al. 2018), it may also drastically increase unemployment. McKinsey estimates that only 5% of jobs can be completely automated using existing technologies, but that 60% of jobs consist of activities that are at least 30% automatable (Manyika and Bughin 2018). Automating part of a job means that fewer humans are required to perform that job, and these estimates suggest that up to 30% of workers, or 800 million people, could be displaced by automation by 2030 (Manyika and Bughin 2018). Such estimates are of course uncertain, but it seems clear that as AI's human-likeness continues increasing, the threat to human jobs increases as well. The economic consequences of massive AI-driven unemployment are therefore potentially profound (Furman and Seamans 2018).

Promising solutions to this issue are scarce. One commonly discussed idea is to tax the corporate beneficiaries of automation and use the resulting revenue to pay every citizen a basic income, thus protecting them from the economic consequences of unemployment (Hughes 2014). However, the political and economic viability of this idea remains unclear, especially in highly capitalistic countries such as the United States (Hoynes and Rothstein 2019). Furthermore, a basic income would do little to address the widening income inequality that is also expected to result from technological unemployment (Berg, Buffie, and Zanna 2016; Bughin et al. 2018), nor would it necessarily alleviate the potential psychological consequences arising from technological unemployment, such as the sense of meaning obtained through work (discussed in more detail below).

These economic consequences of AI adoption are likely to spill over into the political domain as well. A major political trend in recent years has been a shift towards populism, characterized by an anti-establishment and anti-elite orientation, opposition to liberal economics and globalization, xenophobia, and authoritarianism, and reflected by events such as Donald Trump's election, Brexit, and the rise of far-right nationalist political parties in Europe (Inglehart and Norris 2017; Rodrik 2018). This trend is partly fueled by stagnant or worsening economic prospects among the lower- and middle-classes in rich countries coupled with vast increases in wealth among the high upper-class (Inglehart and Norris 2017; Rodrik 2018). AI has the potential to exacerbate this trend by increasing unemployment – especially among unskilled laborers – while further enriching the owners of the technology (Acemoglu and Restrepo 2018), thus threatening to potentially accelerate the populist trend.

Psychological consequences

There are also likely to be psychological consequences to widespread AI adoption. One such consequence that could act as a potentially countervailing force against the populism-increasing technological unemployment is the idea that the increasing salience of AI can reduce intergroup prejudice by presenting a common threat to humanity, in turn making differences among groups of humans seem less important. Indeed, research has found that increasing the salience of AI even in subtle ways can decrease ethnocentrism and increase comfort with racial, religious, and sexual minorities (Jackson, Gray, and Castelo 2019). However, while AI will very likely exacerbate economic conditions among the working class, consumers may attribute those worsening conditions either to technology or to other human groups. Whether the rise of AI ultimately increases or decreases prejudice and populism may depend on whether consumers attribute economic challenges to human groups (i.e., immigrants and foreign workers) or to technology.

An additional psychological consequence already alluded to is the sense of meaningfulness that people often obtain through their work. Meaningfulness refers to the amount of positively valenced significance something holds for an individual (Pratt and Ashforth 2003; Rosso, Dekas, and Wrzesniewski 2010). It has long been known that being employed provides a sense of meaningfulness in part by providing employees with an identity (i.e., a teacher, a scientist, a lawyer, etc.), such that most people would still want to work in order to maintain that identity even if there was no financial incentive to do so (Morse and Weiss 1955; Rosso et al. 2010). Thus, even if a universal basic income were able to provide financial security to those affected by technological unemployment, fewer opportunities to work would also be expected to have negative consequences for people's sense of identity and thereby their sense of meaningfulness.

Any negative effects of widespread AI adoption on people's sense of meaningfulness may be also be driven by a decreased sense of competence and autonomy – two fundamental sources of meaning (Deci and Ryan 2000). As AI outperforms humans at a rapidly growing number of tasks, people's sense of competence may be threatened. Similarly, as AI becomes more widely used to perform tasks previously done by humans, people may have fewer choices regarding how they earn their living, thus threatening a sense of autonomy.

Widespread AI adoption may undermine perceived autonomy in another way as well. Machine learning algorithms are increasingly capable of predicting consumers' desires, identities, and choices based on their online data and then using those predictions to present micro-targeted recommendations (André et al. 2018). Such algorithms can already predict consumers' race, gender, sexual orientation, and personality traits more accurately than the consumers' close friends, based only on publicly available data (Kosinski, Stillwell, and Graepel 2013; Youyou, Kosinski, and Stillwell 2015). Algorithms can also use neuroscientific data to predict a person's decision before they become consciously aware of even having made a decision (Soon et al. 2013). As these algorithms continue to improve, such predictions will become more accurate, with the potential to challenge consumers' sense of autonomy in a deeper sense – potentially including their sense of free will (André et al. 2018).

A more pragmatic consequence may involve the degradation of automated skills. The idea that automating a task normally done by humans will impair those humans' ability to perform those tasks unaided is intuitive. For example, some research suggests that airplane pilots have become less skilled as they rely more on automated pilots (Carr, 2015). Similarly, the advent of online search engines may be impairing memory, as the expectation that one can immediately find answers online reduces the perceived need to store information in one's memory (Sparrow, Liu, and Wegner 2011). As individuals, firms, and governments automate a growing number of important tasks, our society may therefore become dangerously dependent on AI technologies to perform those tasks. This danger highlights the need for at least some subset of humans to continue practicing and maintaining the skills that can be automated, so that they can continue performing the tasks requiring those skills in the event that the technologies fail.

7. CONCLUSION

The potential economic, political, and psychological consequences of AI adoption described above provide a sense of the importance of understanding how consumers and firms adopt AI. These technologies promise to transform society in ways at least as profound as the Internet has already done. Marketers have an important role to play in shaping this transformation by influencing the rate and pattern of AI adoption. This Essay provides a starting point for understanding the context and levers of this powerful influence.

The first contribution this Essay makes is to provide a non-technical description of the specific technologies that underlie AI. This helps to cut through the hype and confusion surrounding that often accompanies promising and complex new technologies without a precise definition. As both marketing practitioners and scholars seek to better understand what AI is and what it can do, this focus on AI's constituent technologies will help.

The second contribution is to provide a significant update over existing and outdated models of technology adoption, which addresses several key limitations of those models and is tailored specifically to AI. The model proposes that in addition to the perceived usefulness of a technology, consumers' comfort with the technology is a key force in shaping adoption, which may be less true in the context of other technologies that lack human-likeness, are less risky, and are less transformative. The model also identifies three broad drivers of both usefulness and comfort and illustrates several components of each of these three drivers: human-likeness, task dimensions, and consumer heterogeneity. This model should also be useful for both practitioners and scholars seeking to understand and influence the process of AI adoption as well as identify new research questions.

The dimensional description of human-likeness is itself a third major contribution. It brings together diverse literatures on mind perception, person perception, dehumanization, and human-robot interaction in order to identify a three-dimensional structure to the concept of human-likeness, consisting of physical, cognitive, and emotional characteristics. Furthermore, the model also applies this structure to better understand what AI is and can do, and how the three dimensions influence the adoption process.

In summary, AI is a collection of five fundamental technologies that provide it with increasing levels of physical, cognitive, and emotional human-likeness, in turn allowing it to perform a wide range of tasks. As such, it offers enormous potential for consumers and firms to obtain services they value through artificial rather than human agents, while also presenting a number of profound potential consequences for the individuals and societies that broadly adopt it. A combination of rigorous research into the causes and consequences of AI adoption, coupled with thoughtful translation of that research into marketing practice, can help tilt the balance away from the negative consequences of AI adoption and towards increased value.

The following two Essays will test several parts of the model proposed here. Essay 2 explores how task dimensions impact consumers' willingness to rely on algorithms, focusing specifically on the perceived objectiveness of the task for which the algorithm is used. This Essay also tests whether the algorithm's affective human-likeness interacts with this task dimension in shaping reliance on algorithms. Comfort and usefulness are measured as mediators throughout this Essay. Essay 3 will then explore how the cognitive, affective, and physical human-likeness of robots impacts consumers' comfort with and perceived usefulness of robots in consumption settings.

CHAPTER 3: ADOPTION OF ALGORITHMS

Algorithms — a set of steps that a computer can follow to perform a task — increasingly outperform humans at many tasks. Pioneering literature from the 1950's demonstrated that very simple algorithms such as linear regression could outperform expert humans on tasks such as diagnosing medical and psychological illnesses (Dawes, Faust, and Meehl 1989; William M. Grove et al. 2000; Meehl 1954). Since then, rapid progress in AI has endowed algorithms with the abilities to understand and produce natural language, learn from experience, and even understand and mimic human emotions. Today, algorithms can outperform even expert humans at an increasingly comprehensive list of tasks, from diagnosing some complex diseases (Simonite 2014) to driving cars and providing legal advice (Krasnianski 2015). Algorithms can also perform seemingly subjective tasks such as detecting emotion in facial expressions and tone of voice (Kodra et al. 2013). Algorithms thus offer enormous potential for improving outcomes for consumers and firms, including the automation of a large proportion of marketing decisions (Bucklin, Lehmann, and Little 1998). The rise of algorithms means that consumers are increasingly presented with a novel choice: should they rely more on humans or on algorithms? Research suggests that the default option in this choice is to rely on humans, even when doing so results in objectively worse outcomes.

ALGORITHM AVERSION

Table 2 summarizes the primary results of empirical studies that have examined trust in and use of algorithms compared to humans, ordered by publication date. The dominant theme is that consumers prefer humans over algorithms (but see Logg, Minson, and Moore 2019 for an exception). For example, people prefer to rely on humans for forecasting student performance after seeing an algorithm err, even when doing so results in suboptimal forecasts (Dietvorst, Simmons, and Massey 2014). People also trust medical recommendations less when they come from an algorithm than from a human doctor (Promberger and Baron 2006). The authors argued that patients prefer to shift responsibility for consequential decisions to someone else, and that it is easier to shift responsibility to humans than to computers. In the same domain, Shaffer et al. (2013) found participants rated physicians who made an unaided diagnosis significantly more positively than a physician who used an algorithm to assist with the diagnosis, but no differently than a physician who consulted a colleague to assist with the diagnosis. Participants who had a greater internal locus of control had more negative evaluations of algorithmic diagnoses.

Önkal et al. (2009) found that participants relied less on advice from an algorithm than from a human when forecasting stock prices. Like Shaffer et al. (2013), they argued that participants find it easier to shift responsibility or blame to other humans. They additionally noted that humans, unlike most algorithms (Armstrong 1980), can provide explanations for their decisions, are seen to have high confidence (Sniezek and Buckley 1995), have a reputation to maintain (Eisenhardt 1989), and have information about future events (Blattberg and Hoch 1990). In contrast, algorithms are thought to possess none of these qualities.

In the domain of employee selection and hiring decisions, Diab et al. (2011) found that participants thought that human interviews were more useful, professional, fair, personal, flexible, and precise than algorithms. In the domain of student performance forecasting, Dietvorst, Simmons, and Massey (2014) found that participants prefered to make their own forecasts rather than relying on an algorithm after seeing the algorithm err, and that while algorithms were seen as better than humans at avoiding obvious mistakes, appropriately weighing attributes, and consistently weighing information, humans were seen as better than algorithms at learning from mistakes, getting better with practice, finding diamonds in the rough, and detecting exceptions to the rule. Dietvorst, Simmons, and Massey (2016) also found that allowing participants to slightly modify the output of an algorithm makes them feel more satisfied with the forecasting process, more tolerant of errors, more likely to believe that the algorithm is superior, and more likely to choose an algorithm to make subsequent forecasts.

Finally, Yeomans et al. (2019) found that participants relied less on an algorithm than on humans for the task of predicting joke funniness, while Logg et al. (2019) found that participants relied *more* on algorithms than on humans for numerical tasks with an objectively correct answer, suggesting that reliance on algorithms varies significantly depending on the type of task for which the algorithm is being used.

Paper	Main Independent Variable	Main Dependent Variables	Main Findings
Promberger and Baron 2006.	Medical recommendation from "physician" vs. "computer program"	Acceptance of recommendation and trust in recommender.	Acceptance and trust are higher for humans vs. computer programs.
Önkal et al. 2009.	Financial forecasts from "human expert" vs. "statistical forecasting method."	Weight of advice (how much participants adjusted own forecast after receiving advice).	Human advice was given more weight.
Diab et al. 2011.	Employee selection via "thorough discussion" vs. "mathematical formula."	Perceived usefulness, fairness, and flexibility of selection method.	Thorough discussion seen as more useful, fair, and flexible than formulae.
Eastwood, Snook, and Luther 2012.	Financial and medical advice based on "intuition and personal experience" vs. "a statistical formula."	Preference, accuracy, fairness, ethicalness of advice methods.	Intuition and experience were preferred and seen as more accurate, ethical, and fair than formulae.
Shaffer et al. 2013.	Doctor who makes an unaided diagnosis or solicits aid from either "computer program" or from "specialist" human colleague.	Doctor's perceived diagnosis ability, professionalism, thoroughness.	Soliciting aid from computer but not from human decreases perceived ability, professionalism, and thoroughness.
Dietvorst, Simmons, and Massey 2014.	Observing vs. not observing an algorithm perform (and err) at forecasting tasks.	Choice to rely on algorithm vs. oneself or algorithm vs. other participants when making incentivized forecasts.	Reduced reliance on algorithm after seeing it err, even when it outperforms humans.
Dietvorst, Simmons, and Massey 2016.	Being able vs. unable to modify an algorithm's forecasts.	Choice to rely on algorithm vs. oneself when making incentivized forecasts.	Increased reliance on algorithm when its output is modifiable.
Logg, Minson, and Moore 2019.	Various forecasts from either "another person" or an "algorithm"	Weight of advice (how much participants	Non-experts rely *more* on advice from algorithms than from

Table 2. Summaries of relevant research on perceptions and use of algorithms.

There are two notable gaps in this literature. First, there has not been a systematic exploration of how and why consumers' willingness to use algorithms varies across the many different types of tasks for which algorithms can be used. Second, there has been little exploration and validation of practical interventions that marketers can use to increase consumers' willingness to rely on algorithms instead of humans, especially in cases where the algorithm outperforms expert humans.

I therefore make two primary contributions in this Essay by addressing these gaps. First, I examine how willingness to trust and use algorithms varies by characteristics of the task. I identify a robust effect that algorithms are trusted and used less for tasks that are seen as subjective in nature and show that this effect occurs primarily because of a belief that algorithms are not useful for subjective tasks. Trust involves both cognitive and affective dimensions (Johnson and Grayson 2005). Cognitive trust involves confidence in another agent's performance or reliability, while affective trust is based on one's feelings and can be independent from performance. In the context of this research, I suggest that consumers' overall trust in algorithms is affected by both performance-based, cognitive beliefs about the algorithm's performance, as well as by feelings stemming from consumers' comfort with the use of algorithms for tasks normally done by humans, which can be independent of performance-related beliefs. I therefore explore both consumers' beliefs about algorithmic performance as well as their comfort with the use of algorithms as mechanisms of this main effect.

Second, I explore approaches for making algorithms more attractive to potential users when use is low despite algorithmic superiority over expert humans. I show that the perceived objectiveness of a task is malleable and that re-framing tasks as being relatively objective increases trust in and willingness to rely on algorithms. Furthermore, I show that the belief in algorithm usefulness for subjective tasks is itself also malleable. Specifically, increasing the perceived *affective* human-likeness of algorithms by providing real examples of algorithms with affective abilities, such as understanding emotion and creating art, can make algorithms seem more useful at performing subjective tasks, which ultimately increases reliance on algorithms for such tasks.

HYPOTHESIS DEVELOPMENT

There are many potentially relevant dimensions along which tasks vary that can impact consumers' use of algorithms. For example, consumers are already familiar with the use of algorithms for certain tasks such as recommending movies on Netflix or filing taxes on TurboTax. Familiarity with algorithms for a given task is likely to increase trust in and willingness to rely on algorithms for that task. Similarly, some tasks are much more consequential than others, in the sense that performing the task poorly will have more serious consequences. Consumers may be less willing to trust and rely on algorithms for more consequential tasks because doing so poses greater risks. More theoretically relevant, however, is the perceived *objectiveness* of the task. I define an objective task as one that involves facts that are quantifiable and measurable, compared to subjective tasks, which I define as being open to interpretation and based on personal opinion or intuition. Research has shown that lay people see objective tasks as requiring logical, rule-based analysis, and subjective tasks as requiring intuition and "gut instincts" (Inbar et al. 2010). Importantly, the objectiveness of a task is not completely inherent in a given task but may be a malleable perception with heterogeneity both among different people and over time. I will therefore exploit this heterogeneity in order to develop manipulations of perceived task objectiveness.

The impact of task objectiveness on consumers' trust and use of algorithms likely depends on what kind of abilities consumers typically believe that algorithms possess. One major conceptual distinction that is relevant here is between cognitive and emotional abilities. For example, research on dehumanization has shown that people perceive two categories of human abilities. First are "human uniqueness abilities," which distinguish humans from other animals but can be shared with machines. These tend to be cognitive in nature (such as logic and rationality). Second are "human nature abilities," which may be shared with other animals but not with machines. These tend to be affective or emotional in nature (such as warmth and intuition; Haslam 2006; Loughnan and Haslam 2007). Importantly, research has shown that machines such as robots are seen as lacking human nature abilities (which are emotional) but not human uniqueness abilities (which are cognitive; Haslam et al. 2008).

Research on mind perception has focused on two similar dimensions: agency, the ability to engage in intentional planning and action, and experience, the ability to subjectively experience emotions and sensations. Mirroring research on humanness, machines (such as robots) are seen as having some degree of agency but no experience (Gray et al. 2007), and endowing robots with experience creates more negative reactions than endowing robots with agency (Gray and Wegner 2012). These streams of research demonstrate that consumers perceive human abilities as either cognitive or emotional and are willing to grant machines more cognitive than emotional abilities. Integrating these streams of research with the distinction between objective tasks, which are typically associated with more "cognitive" abilities, and subjective tasks, which are typically associated with more "emotional" abilities (Inbar, Cone, and

Gilovich 2010), suggests that consumers will believe that algorithms will be less useful for subjective tasks because they are believed to lack the affective or emotional abilities typically associated with such tasks. Beliefs about whether or not a technology is useful are fundamental determinants of whether that technology is ultimately adopted (Davis et al. 1989; Rogers 1976). My first hypothesis is therefore:

H1: Consumers will trust and rely on algorithms less for subjective (vs. objective) tasks.

I measure trust in algorithms in several studies because research has shown that trust in a technology is an important determinant in the decision to use it (Komiak and Benbasat 2006; Li, Hess, and Valacich 2008; Pavlou and Gefen 2004). Trust is relevant in situations where one party is somehow dependent on the actions of another party and this dependence involves risk (Chopra and Wallace 2003), which is the case in situations where consumers use an algorithm to perform a task normally done by a human. Higher trust in an algorithm should therefore lead to greater willingness to use the algorithm. Thus, while the majority of the studies focus specifically on the actual or intended use of algorithms, I also measure trust in algorithms (both in terms of affective and cognitive trust) as an important factor that contributes to use.

It follows from H1 that one way of increasing the use of algorithms for a given task is to increase the degree to which the task is seen as being objective. Most tasks can be seen as more or less objective depending on how the task is framed and which components of the task are emphasized. Specifically, a given task can be approached either by measuring and analyzing relevant quantitative variables, or by using intuition or gut feelings. For example, evaluating and hiring a job candidate could be done by using standardized psychometric tests and measures, or by conducting informal interviews and relying on one's gut feeling or intuition. Importantly, it is not always clear which of these approaches is superior for many important tasks, as psychologists continue to debate the relative merits of more "deliberate" vs. more "intuitive" approaches to different tasks (Bear and Rand 2016; Dijksterhuis et al. 2006; Gigerenzer and Brighton 2009; Slovic et al. 2004). Furthermore, consumers also differ in terms of their preferences for and tendencies to rely on more analytical vs. intuitive approaches to decision making (Greifeneder, Bless, and Pham 2011; Inbar et al. 2010). This uncertainty provides an opportunity to frame tasks normally seen as subjective as being more objective. The second hypothesis is therefore:

H2: Describing a task as benefiting from quantitative analysis (relative to intuition) will increase perceived task objectiveness and consumers' trust in and reliance on algorithms.

In addition to changing how the task is perceived, a second approach for increasing the use of algorithms involves changing how the algorithm itself is perceived. As mentioned earlier, consumers tend to believe that machines lack fundamentally human capabilities that are emotional or affective in nature (i.e., that they lack *affective* human-likeness) (Gray et al. 2007; Haslam et al. 2008). However, this belief is increasingly inaccurate. Algorithms can already create paintings that sell for hundreds of thousands of dollars (Quackenbush 2018), write compelling poetry and music (Deahl 2018; Gibbs 2016b), predict which songs will be hits (Herremans, Martens, and Sörensen 2014), and even accurately identify human emotion from facial expressions and tone of voice and respond accordingly (Goasduff 2017; Kodra et al. 2013; McDuff et al. 2013). Algorithms are therefore increasingly capable of performing the kinds of tasks typically associated with subjectivity and emotion. Note that even though algorithms may accomplish these tasks using very different means than humans do – i.e., using pre-determined computer programming rather than intuitions or gut feelings – I will show that the fact that algorithms can accomplish such tasks at all makes algorithms seem more human-like.

I expect that increasing algorithms' perceived human-likeness in this way will moderate the effect of task objectiveness. This moderation, however, could plausibly either increase or decrease the effect of objectiveness. On one hand, increasing affective human-likeness is likely to increase the perceived usefulness of algorithms for subjective tasks, since consumers believe that subjective tasks require affective abilities that algorithms are normally thought to lack. Making algorithms seem more human-like could therefore decrease or eliminate the main effect of task objectiveness. This result would indicate that cognitive trust (i.e., beliefs about algorithm usefulness) is more important than affective trust (i.e., feelings that are independent from beliefs about usefulness) in shaping consumers' use of algorithms.

On the other hand, increasing affective human-likeness may also produce discomfort with the use of algorithms by challenging the belief that humans are distinct from machines. Indeed, social identity theory posits that people derive meaning and satisfaction from membership in distinct groups, and that when an outgroup threatens the sense of distinctiveness of their in-group, they react negatively towards the threatening outgroup (Tajfel 1982). In other words, people like to believe that their ingroup is unique, and when an outgroup begins to challenge that perceived uniqueness, the outgroup will be evaluated negatively (Brewer 1991; Ferrari, Paladino, and Jetten 2016). Increasing the affective human-likeness of algorithms could therefore represent an intergroup challenge in the sense of algorithms as an outgroup challenging the distinctiveness of humans (as an ingroup) from machines. According to social identity theory, this challenge to ingroup distinctiveness could in turn lead to negative evaluations of the challenging outgroup (algorithms in this case), ultimately decreasing the use of algorithms including for subjective tasks. This result would suggest that affective trust is more important than cognitive trust in determining consumers' use of algorithms.

Measuring consumers' reliance on algorithms that vary in affective human-likeness can thus help to tease apart these two competing hypotheses. Whether increasing affective human-likeness reduces or exacerbates the main effect of task objectiveness may ultimately depend on whether cognitive factors (i.e., beliefs about usefulness) or affective factors (i.e., feelings of discomfort potentially stemming from intergroup challenges) have a stronger impact on consumers' use of algorithms. In place of a third hypothesis, I thus posit a final research question:

Will increasing algorithms' perceived human-likeness reduce or increase the effect of perceived task objectiveness on consumers' use of algorithms?

I test these hypotheses and answer this research question using a variety of field and lab studies and several different dependent measures. In order to increase generalizability and demonstrate the robustness of the effects, I operationalize the dependent variable in multiple ways, including self-reported trust in and preference for algorithms relative to humans, clicks on online advertisements for algorithm- and human-based services, and actual reliance on algorithms in the context of an incentivized task.

To summarize, existing explanations of algorithm aversion suggest that it is largely driven by a perception that algorithms lack human abilities. I propose that it is specifically affective abilities that algorithms are seen as lacking and that this belief decreases willingness to rely on algorithms for tasks that seem subjective. Consequently, emphasizing that a given task benefits from a more quantitative approach can increase the perceived objectiveness of that task, ultimately increasing trust and use of algorithms. Finally, increasing algorithms' perceived affective human-likeness may either reduce or increase the effect of perceived task objectiveness on consumers' use of algorithms, depending on whether algorithmic usefulness or discomfort with human-like algorithms have a stronger impact on use.

I explore these questions in 6 studies. Study 1 shows that trust in algorithms varies substantially depending on the task, and that trust is lower for more subjective tasks. Study 2 replicates this effect in an applied field study. Study 3 shows that providing evidence that algorithms are superior to humans for a specific task is less useful at increasing consumers' preference for using algorithms when the task is relatively subjective. Study 4 shows that re-framing a subjective task as being more objective increases trust in algorithms. Study 5 replicates the task-framing effect in a field study. Finally, Study 6 shows that increasing algorithms' perceived human-likeness by providing examples of algorithms with human nature abilities *increases* the use of algorithms for subjective tasks, thus eliminating the effect of task objectiveness. Put simply: consumers have strong preconceptions about what algorithms are good at, so two ways to increase reliance on algorithms are to change the way the task is perceived and change the way the algorithm's abilities are perceived.

Figure 2 depicts the conceptual model that I test in this Essay. The main effect that I demonstrate is that perceived task subjectivity reduces consumers' trust in and willingness to rely on algorithms (Studies 1–6). I provide evidence that this effect is explained partially by the perceived usefulness of algorithms for subjective tasks (Studies 4 and 6), and partially by consumers' discomfort using algorithms for subjective tasks (Study 6). Furthermore, I study the effects of algorithms' affective human-likeness, finding both direct effects on discomfort and interactions between human-likeness and task objectiveness in shaping discomfort, perceived usefulness of the algorithm, and reliance on the algorithm (Study 6).

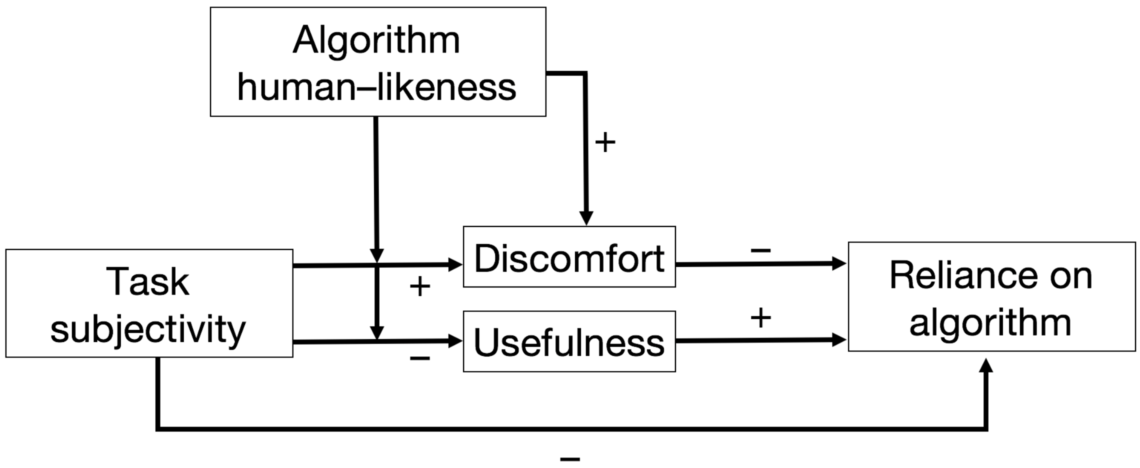


Figure 2. Conceptual Model Tested in Essay 2.

STUDY 1

To gain an initial understanding of how trust in algorithms varies by task, I examined 26 different tasks that vary along several dimensions. The primary goal of this study was to test H1, that trust in algorithms would be lower for more subjective tasks. I also measured task consequentialness and how familiar participants were with the use of algorithms for each task, two other potentially relevant dimensions. Furthermore, I also measured trust in well-qualified humans for each task so that I could compare trust in algorithms to trust in humans for each task. This study was conducted in two parts, with one sample of participants rating the tasks along the dimensions of objectiveness, consequentialness, and familiarity with the use of algorithms, and a second sample rating their trust in algorithms or in humans for each task.

Method

Participants and design. For part 1, I recruited 250 participants (M_{age} = 37, 41% female) from Mechanical Turk (MTurk), who rated tasks along several dimensions. For part 2, I recruited 387 participants (M_{age} = 36, 45% female) from MTurk, who were randomly assigned to one of two conditions (trust in humans vs. trust in algorithms).

Procedure. For part 1, participants rated each of 26 tasks on how objective vs. subjective it seemed, how consequential vs. inconsequential it seemed, and how familiar they were with the use of algorithms for each task, using scales from 0 (not at all) to 100 (completely). The tasks as well as the dimensions being rated were presented in random order. For part 2, participants indicated how much they would trust *either* an algorithm or a "very well qualified person" for each of 26 tasks also on a scale from 0 (not at all) to 100 (completely). For example, for the task of diagnosing a disease, the person was described as a doctor. The tasks are shown in Table 3.

Results and Discussion

Averaged across tasks, trust in a qualified human was higher than trust in algorithms (M_{human} = 70.2 vs. $M_{algorithm}$ = 52.8, $t(385) = 5.75$, $p < .001$). However, trust in algorithms was higher than in humans for certain tasks (predicting stock market outcomes, predicting the weather, analyzing data, and giving directions, t's > 4.70, p's < .001; see Table 3).

	Trust Human	Trust Algorithm	Human–Algorithm Gap	Task Objectiveness
Predicting Joke Funniness	65	30	**35**	27
Hiring & Firing Employees	72	34	**38**	49
Rec. Romantic Partner	59	37	**22**	26
Writing News Article	79	37	**42**	48
Pred. Recidivism	54	42	**12**	45
Composing a Song	81	43	**38**	30
Driving Truck	81	43	**38**	70
Rec. Gift	75	46	**29**	26
Pred. Student Performance	63	46	**17**	52
Piloting Plane	79	47	**32**	78
Driving Car	81	47	**34**	69
Rec. Disease Treatment	73	48	**25**	69
Disease Diagnosis	73	48	**25**	77
Pred. Employee Performance	61	50	**11**	51
Driving Subway	77	52	**25**	73
Pred. Election	51	54	-3	57
Rec. Marketing Strategy	70	56	**14**	55
Recommending Music	75	59	**16**	22
Rec. Movie	76	59	**17**	23
Buying Stocks	62	60	2	56
Playing Piano	84	61	**23**	48
Pred. Stocks	55	63	**-8**	58
Pred. Weather	57	67	**-10**	68
Scheduling Events	78	69	**9**	62
Data Analysis	69	80	**-11**	73
Giving Directions	70	82	**-12**	75

Table 3. Consumers' trust in algorithms vs. qualified humans. Note: Tasks are listed in increasing order of trust in algorithms. The human-algorithm gap for a task is statistically significant ($p < .001$) when the corresponding number is in bold.

In order to test the effects of the three dimensions of consequentialness, familiarity, and objectiveness on trust in algorithms, I conducted a regression in which these three dimensions were simultaneously used to predict trust in algorithms. This revealed that trust in algorithms was lower for tasks that seemed more consequential ($\beta = -.56$, $p < .001$), higher for tasks for which consumers were more familiar with the use of algorithms ($\beta = .42$, $p < .001$), and most importantly higher for tasks that seemed more objective ($\beta = .46$, $p = .004$). The adjusted R^2 for this regression was .54. These results provide initial support for H1, suggesting that trust is higher for more objective tasks. The next study will corroborate this finding in a field study.

STUDY 2

While the first study provided initial support for H1, it is possible that participants' self-reported trust is not a reliable indicator of their actual behavior. In order to address this concern, I examined the role of task objectiveness in a field study in which participants' behavior was directly observed.

Method

Participants and design. I created 4 advertisements, organized in a 2 (human vs. algorithmic advisor) x 2 (dating vs. financial advice) design, and displayed them on Facebook (see Appendix A for the exact stimuli used in this and all other studies). The ads portrayed either a human or an algorithm providing either dating advice (rated as highly subjective in Study 1) or financial advice (rated as highly objective in Study 1). These 4 ads were seen by 41,592 unique Facebook users (40% female, mean age not observed) on their Facebook Newsfeed (i.e. the stream of posts that users see on Facebook from their friends and advertisers).

Procedure. Participants who see ads on their Facebook Newsfeed can click on those ads to learn more about them. The ads were clicked on 604 times in total. Participants who clicked on the ads were taken to a page explaining that I was studying consumers' trust in algorithms. The dependent variable was the click-through rate (CTR) of the ads, which is the number of clicks a given ad received divided by the number of times that ad was seen (Facebook shares this information with the creator of the advertising campaign). I expected participants to be more likely to click on an ad for dating advice when it was advertised as coming from a human vs. an algorithmic advisor, but that click-through rates would not differ between human and algorithmic financial advice, because the latter is a more objective task.

Results and Discussion

I conducted a logistic regression to estimate the effects of human vs. algorithm (human = 1, algorithm = 0), finance vs. dating ad (finance = 1, dating = 0), and their interaction on the CTR (click = 1, no click = 0). This revealed that the CTR was higher for human ads ($\beta = 1.14$, $p < .001$) and finance ads ($\beta = .59$, $p = .001$). I also found a significant interaction between these factors ($\beta = -.99$, $p < .001$). As predicted, the click-through rate for the dating advice ads was significantly higher for the human advisor (2.1%) than for the algorithm advisor (0.6%, $\chi^2(1) = 29.10$, $p < .001$). In contrast, for the financial advice ads, the click-through rate was only marginally significantly higher for the human advisor (1.8%) than for the algorithm advisor (1.6%, $\chi^2(1) = 3.26$, $p = .071$; see Figure 3). This range of CTRs is comparable to other recent studies using Facebook advertising campaigns (Matz et al. 2017). These results provide further support and external validity to the notion that trust in algorithms is low primarily for tasks that are seen as subjective.

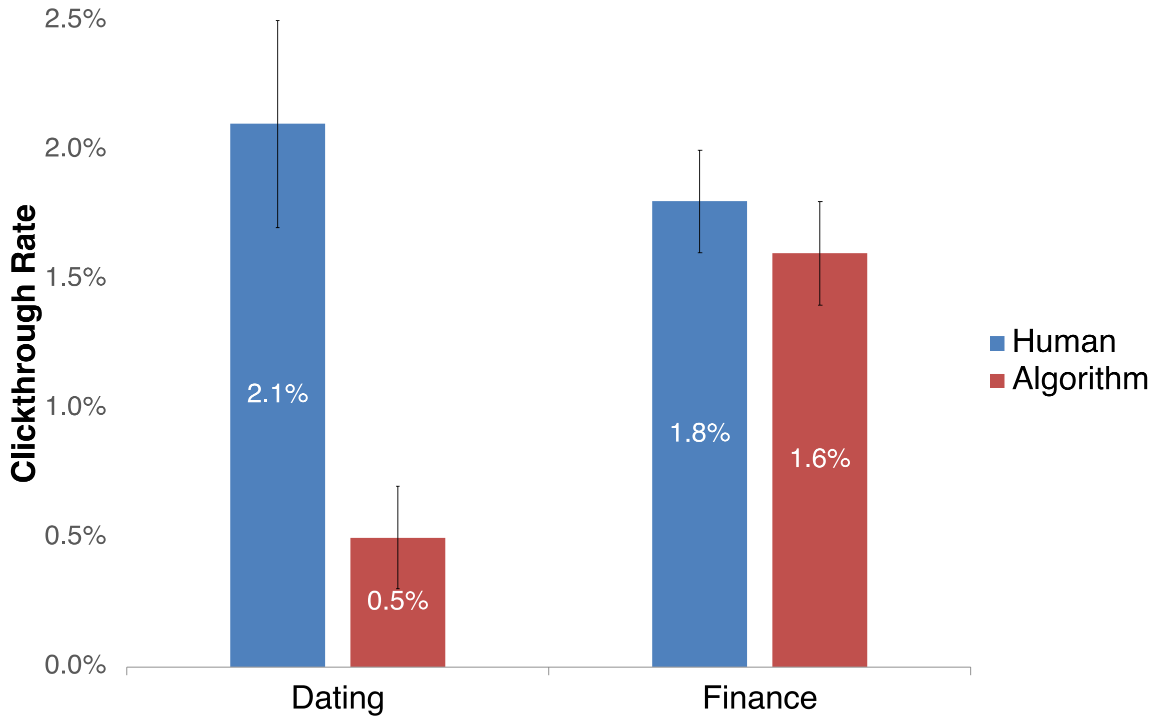


Figure 3. Trust in humans vs. algorithms is higher for a subjective but not objective task. Note: Error bars represent standard errors.

STUDY 3

Our first two studies have provided support for H1, showing that consumers are less willing to trust and use algorithms for tasks that seem subjective. Nevertheless, algorithms are often highly effective at such tasks. Interventions that can increase trust and use of algorithms for such tasks would therefore be helpful to both consumers and firms. One of the most intuitive approaches for increasing consumers' willingness to use algorithms is to provide them with empirical evidence of the algorithms' superior performance relative to humans for the specific task in question. However, given the effect demonstrated in the previous studies, I anticipated that this evidence would be less effective at increasing willingness to use algorithms for tasks that seem subjective because consumers may be less likely to believe that algorithms can perform subjective tasks better than humans even when provided evidence to the contrary. Note that this study does not manipulate the perceived human-likeness of the algorithm, which itself may impact the perceived usefulness of algorithms in general, but instead manipulates the perceived usefulness of the algorithm for the specific task in question.

Method

Participants and design. 201 MTurk users (M_{age} = 36, 49% female) reported their preference for using an algorithm relative to a qualified human for 9 tasks (see Table 4). These 9 tasks varied in terms of both trust in algorithms and perceived objectiveness as measured in Study 1. Importantly, there is published research available documenting the superiority of algorithms over qualified humans for all 9 tasks. Participants were assigned to one of two performance conditions: known performance or unknown performance. In the known performance condition, I told participants that the algorithm outperformed the human and described the results of a real study that had demonstrated the algorithm's superior performance. In the unknown performance condition, this information was omitted, and the performance of the algorithm was not mentioned.

Procedure. Participants read about and rated each of the task 9 tasks individually. In the known performance condition, participants read about a published academic study for each task which demonstrated that an algorithm could outperform qualified humans. I provided links to each study and reported how much better the algorithm performed compared to the humans in the study. In the unknown performance condition, this information was not provided, and participants simply reported their preference without learning how the performance of algorithms compared to the performance of humans.

Participants reported whether they would rather use an algorithm or a qualified human for each task, with responses entered on a 0 to 100 scale, with 0 labeled as the relevant qualified human, 50 labeled as no preference, and 100 labeled as algorithm. For example, the relevant qualified human was "human doctor" for the task of diagnosing a disease and "human judge" for the task of deciding a parole case.

Results and Discussion

Preference for using algorithms was higher when performance data was provided (M = 50.4) compared to when it was not (M = 28.6, $t(199) = 17.5$, $p < .001$). The effect of performance data was highly significant for each of the 9 tasks (see Table 4). I assigned each task the objectiveness score that it received in Study 1, and then conducted an ANOVA in which performance condition, task objectiveness, and their interaction were used to predict preference for using an algorithm relative to a human. This revealed main effects for performance condition, $F(1,183) = 208.23$, $p < .001$, task objectiveness, $F(8,183) = 4.94$, $p = .026$, and a significant interaction, $F(8,183) = 6.12$, $p = .013$. The effect of providing performance information was significant for each task. In order to explore the interaction, I divided tasks into "objective tasks" (rated as greater than 50, or the midpoint used to measure objectiveness), and "subjective tasks" (rated as less than 50). The effect of providing performance data on preference for using an algorithm was significantly greater for objective tasks ($M_{performance\_data} = 55.1$, $M_{no\_performance\_data} = 34.5$, $t(199) = 14.37$, $d = .64$, $p < .001$) than it was for subjective tasks ($M_{performance\_data} = 47.3$, $M_{no\_performance\_data} = 31.7$, $t(199) = 10.35$, $d = .51$, $p < .001$).

	Preference for Algorithm Relative to Human			
	Without performance data	With performance data	Δ	Task objectiveness
Predict Student Performance	**39**	**57**	18	52
Predict Employee Performance	**27**	52	25	51
Recommend Disease Treatment	**31**	**59**	28	69
Predict Recidivism	**24**	52	28	45
Drive Car	**26**	53	27	69
Recommend Movie	**33**	52	19	23
Diagnose Disease	**23**	46	23	77
Predict Personality	**35**	**40**	5	41
Predict Joke Enjoyment	**19**	**35**	16	27
Average	**29**	50	21	50

Table 4. Consumers' trust in humans vs. algorithms with and without performance data provided. Note: higher numbers indicate greater trust in algorithms relative to humans. Bolded means are significantly different from the scale midpoint (50, labeled as "trust both equally"). The next-to-last column is the change in preference between condition (significant for all tasks), and the last column is the tasks' rated objectiveness, taken from Study 1.

The results of this study suggest that consumers' willingness to use an algorithm instead of a qualified human can be increased simply by demonstrating that the algorithm outperforms the human, although the increase is significantly smaller for subjective tasks. This study therefore provides further support for H1 by showing that task subjectivity reduces willingness to use algorithms even when consumers are made explicitly aware that the algorithm outperforms humans. However, participants remained roughly indifferent between humans and superior algorithms even for several of the most objective tasks. Indifference is insufficient for marketers interested in selling algorithm-based products and services. Importantly, this indifference is suboptimal for consumers when algorithms outperform humans (as is the case in each of the 9 tasks in this study), since consumers who are indifferent will tend to select the default or status quo option (i.e., the inferior performing human; Dinner et al. 2011). These findings emphasize the need for other approaches to increase trust and use of algorithms. The next set of studies test H2, which posits that re-framing the task for which an algorithm is used can be one such approach.

STUDY 4

In this study, I attempt to increase consumers' trust in algorithms for subjective tasks by emphasizing that such tasks can benefit from quantitative analysis (relative to intuition), thus providing a test of H2. This study also tests both the cognitive and affective dimensions of trust, confirming that the effects are driven by the cognitive dimension.

Method

Participants and design. 201 Prolific Academic users (M_{age} = 33, 47% female) were randomly assigned to one of two conditions in which two tasks were described in such a way that emphasized either their quantitative components (the *objective* condition) or their intuitive components (the *subjective* condition). The tasks were recommending a movie and recommending a romantic partner, two consumer-relevant tasks that were rated as highly subjective in Study 1. Prolific Academic is a crowdsourcing website where participants are less familiar with common experimental paradigms and more honest than participants on Mechanical Turk (Peer et al. 2017).

Procedure. In the subjective condition, participants read that, according to previous studies, the tasks were best accomplished by focusing on one's moods, emotions, and intuitions. In the objective condition participants read that, according to previous studies, the tasks were best accomplished by focusing on quantifiable data such as measured personality traits (see Appendix A for exact stimuli). Participants reported how much they would trust an algorithm relative to a human for the two tasks. Responses were entered on 0–100 scales, with the scale anchored at 0 (trust human more), 50 (trust both equally), and 100 (trust the algorithm more). The qualified human was specified as a friend for the movie task and a professional matchmaker for the dating task. Participants also reported how objective the tasks seemed.

In order to measure both the affective and cognitive components of consumers' trust in algorithms, I asked participants how much they agreed with the following questions: for cognitive evaluations, "I can see the benefits in algorithms that can perform this kind of task better than humans," "Algorithms that can perform this kind of task could be useful," and "I believe this kind of algorithm can perform well," and for affective evaluations, "Algorithms that can perform this kind of task better than humans make me uncomfortable," "Algorithms that can perform this kind of task go against what I believe computers should be used for," and "Algorithms that can perform this kind of task are unsettling." Alphas were .96 and .92 respectively, and all items were anchored at 0 "not at all" and 100 "completely." I refer to the two kinds of evaluation as "usefulness" and "discomfort" in the following analyses.

Results and Discussion

Participants trusted humans more than algorithms in both conditions for both tasks (t's > 1.89, p's < .062 comparing each mean to 50 [trust both equally]). However, emphasizing the quantitative approach to accomplishing the tasks succeeded at increasing trust in algorithms for both tasks (for movie recommendation: $M_{subjective} = 31.5$, $M_{objective} = 45.1$, $t(199) = 4.04$, $p < .001$, and for romantic partner recommendation: $M_{subjective} = 36.4$, $M_{objective} = 43.8$, $t(199) = 2.09$, $p = .038$). Looked at another way, 36% of participants in the quantitative framing condition reported trusting an algorithm more than a human (> 50 on the scale, collapsing across both tasks), compared to only 19% of participants in the intuitive framing condition ($\chi^2(1) = 6.60$, $p = .010$).

Emphasizing the quantitative approach to the tasks also increased the degree to which the tasks were seen as objective (collapsing across the two tasks: $M_{subjective} = 32.1$, $M_{objective} = 40.7$, $t(199) = 3.11$, $p = .002$) and made algorithms seem more useful for the tasks ($M_{subjective} = 70.7$, $M_{objective} = 76.9$, $t(199) = 1.96$, $p = .052$). The manipulation had no effect, however, on participants' discomfort with the use of algorithms for the tasks (reverse coded: $M_{subjective} = 72.0$, $M_{objective} = 72.1$, $t(199) = 0.27$, $p = .795$). Discomfort with the use of algorithms on its own did have a significant negative effect on trust in the algorithm ($\beta = -.19$, $p < .001$), while perceived usefulness had a significant positive effect on trust ($\beta = .34$, $p < .001$). However, using both discomfort and perceived usefulness to predict trust, usefulness remained a significant predictor ($\beta = .34$, $p < .001$) while discomfort was not ($\beta = -.004$, $p = .948$), suggesting that any initial effect of discomfort can be reduced if the algorithm is seen as being useful.

A mediation analysis with 5,000 bootstrapped replications confirmed that perceived task objectiveness and perceived usefulness of algorithms mediated the relationship between task framing and trust in algorithms. As reported above, task framing affected perceived task objectiveness, which in turn affected the perceived usefulness of algorithms for the tasks (β = .20, p = .007). The direct effect of task framing on trust in algorithms (β = -10.49, $p < .001$) was reduced but still significant when accounting for the mediators (β = -6.04, p = .012), and the indirect effect was significant (β = –.002, 95% CI [–.0002, –.0117]).

These results demonstrate that the perceived objectiveness of a given task is malleable, and that objectiveness impacts both the perceived usefulness of algorithms for a task and self-reported trust in the algorithm for that task. These findings therefore suggest a practical marketing intervention that can be used to increase trust in and use of algorithms for tasks that are typically seen as subjective. I test this intervention using a field study in Study 5.

STUDY 5

In order to increase the external validity of the findings from Study 4 and test whether they can be practically useful for marketers, I conducted a second Facebook advertising study in which I manipulate the perceived objectiveness of subjective tasks in the context of a Facebook advertising campaign for algorithm- and human-based dating services.

Method

Participants and design. I created 2 advertisements for an algorithm-based dating service that either highlighted a quantitative approach to choosing a romantic partner or did not (using the more neutral ad for algorithm-based dating advice from Study 2; see Appendix A for ads). I displayed these ads on Facebook and they were seen by 13,621 Facebook users (39% female, mean age not observed).

Procedure. The ads were clicked on 110 times in total. As in Study 2, participants who clicked on the ads were taken to a page explaining that the researchers were studying consumers' trust in algorithms, and the dependent variable was the click-through rate (CTR) of the ads.

Results and Discussion

Replicating the results of Study 4, framing dating advice as benefiting from a quantitative approach increased the click-through rate (0.87%) compared to the control ad (0.39%, $\chi 2(1) = 3.74$, $p = .053$). Framing a task that is normally seen as highly subjective as in fact benefiting from quantitative data thus provides marketers with a practical tool for increasing consumers' willingness to use algorithm-based products for tasks in which algorithm aversion might otherwise occur.

STUDY 6

Our final study attempts to increase the use of algorithms for subjective tasks in a different way. Instead of providing data regarding the algorithm's performance at the specific task in question or re-framing the task itself as benefiting from quantification, I instead attempt to increase the perceived affective human-likeness of the algorithm by providing real examples of algorithms performing tasks that are typically thought of as requiring emotional and intuitive abilities – i.e., the kinds of abilities that machines are thought to lack and that are seen as necessary for subjective tasks (Gray et al. 2007; Haslam et al. 2008; Inbar et al. 2010). Although social identity theory suggests that this approach may also make consumers less comfortable with the use of algorithms by challenging the belief in human distinctiveness from machines, the results of Study 4 suggest that the perceived usefulness of an algorithm is a stronger influence than discomfort on trust in algorithms. I therefore expect that increasing affective human-likeness will make algorithms seem more useful at subjective tasks and thus increase the use of algorithms for such tasks, despite potentially also creating discomfort with the idea of a human-like algorithm.

Recall that prior research has consistently identified two dimensions of human-likeness, corresponding to cognitive and affective abilities (also called agency and experience or human uniqueness and human nature; Haslam 2006; Gray, Gray, and Wegner 2007). I chose to focus specifically on manipulating the affective dimension of human-likeness in this study because that dimension is the most relevant to subjective tasks, and because that dimension is the one commonly seen as distinguishing humans from machines. Algorithms with affective human-likeness are therefore likely to be seen as both more useful for subjective tasks as well as more threatening to human distinctiveness from machines. This study therefore further helps to determine the relative effects of perceived usefulness (i.e., cognitive trust) and discomfort (i.e., affective trust) on the use of algorithms.

Method

Participants and design. 399 participants from Prolific Academic (49% female, mean age = 35.2) were assigned to one of four conditions in a 2 (affective human-likeness: high vs. low) x

2 (task framing: subjective vs. objective) between-subjects design.

Procedure. In the high human-likeness conditions, participants read that algorithms can perform a range of tasks that are typically thought of as contributing to affective human-likeness, including creating music and art, predicting which songs will be popular, and understanding people's emotions. In the low human-likeness conditions, participants read that algorithms cannot perform these kinds of tasks. In reality, algorithms can in fact perform these tasks. I chose these specific tasks because both creativity and emotional sensitivity are considered to be fundamental components of human nature – i.e., the affective dimension of human-likeness (Haslam et al. 2005). Participants were asked to summarize the information they read in this section to ensure they paid attention to the material. As a manipulation check, participants reported how much they agreed with the statement "Algorithms that can perform this kind of task make humans seem less distinct from machines" on a 0 (not at all) – 100 (completely) scale.

Participants were then shown a graph of the value of the S&P 500 stock market index over the past year and were asked to estimate its value 30 days in the future. Before providing their initial estimate, participants were informed that the 5% most accurate estimates would be rewarded with a bonus payment six times larger than their base compensation, in order to incentivize accuracy and encourage serious engagement with the task.

After making their initial estimate, participants were told that an algorithm designed by an expert financial advisor had also made an estimate, were shown the algorithm's estimate, and were given the opportunity to revise their initial estimate. This paradigm is known as the Judge Advisor System and is commonly used to measure reliance on advice by computing how much participants revise their initial estimate in response to external advice (Logg et al. 2019; Sniezek and Buckley 1995). Reliance on advice is measured as the difference between the final and initial estimates produced by each participant divided by the difference between the advice and the initial estimate.

The algorithm's estimate was accompanied by a manipulation of task objectiveness. Specifically, in the objective framing condition, participants were told that there are clear mathematical relationships between economic measures such as supply and demand and the price of a stock, and that relying on these objective indicators is therefore the best way to estimate a stock's future value. In contrast, in the subjective framing condition, participants were told that human feelings and intuition are the primary drivers of stock prices, and that relying on these subjective factors is therefore the best way to estimate a stock's future value.

Finally, to measure perceived usefulness of the algorithm and discomfort with the use of algorithms, participants were also asked how much they agreed with the following statements: "I believe this kind of algorithm can perform well," and "Algorithms that can perform this kind of task better than humans make me uncomfortable." These items were measured on a 0 (not at all) – 100 (completely) scale.

Results and Discussion

I excluded 33 participants (8.27% of the sample) whose summaries of the human-likeness manipulation clearly indicated that they had not read the information, resulting in a final sample of 366. The results reported here are nearly identical if I include these participants. I first confirmed that the manipulation of human-likeness was effective: participants thought that humans seem less distinct from machines in the high human-likeness condition (M = 45.8) than in the low human-likeness condition (M = 37.8), $t(364) = 2.75$, $p = .006$). I computed reliance on the algorithm's advice by dividing the difference between participants' final and initial estimate by the difference between the advice and their initial estimate. This produces a measure that ranges, in most cases, from 0 (complete discounting of the advice) to 1 (complete reliance on the advice). A value of .30 thus corresponds to a 30% reliance on advice, which is the typical value seen in the advice taking literature (Soll and Larrick 2009).

A 2x2 ANOVA revealed that reliance on the algorithm was significantly impacted by the task framing (objective vs. subjective), $F(1, 363) = 10.01$, $p = .002$, non-significantly by providing information about algorithms' human-likeness, $F(1, 363) = 2.24$, $p = .134$, and marginally by the interaction between these two factors, $F(1, 363) = 3.52$, $p = .060$. When participants were told that algorithms have *low* human-likeness, the effect of task framing was significant, as in prior studies ($M_{subjective} = .22$, $M_{objective} = .39$, $t(364) = 3.30$, $p = .001$, see Figure 4). However, when participants were told that algorithms have *high* human-likeness, the effect of task framing was no longer significant ($M_{subjective} = .35$, $M_{objective} = .40$, $t(364) = 1.03$, $p = .303$).

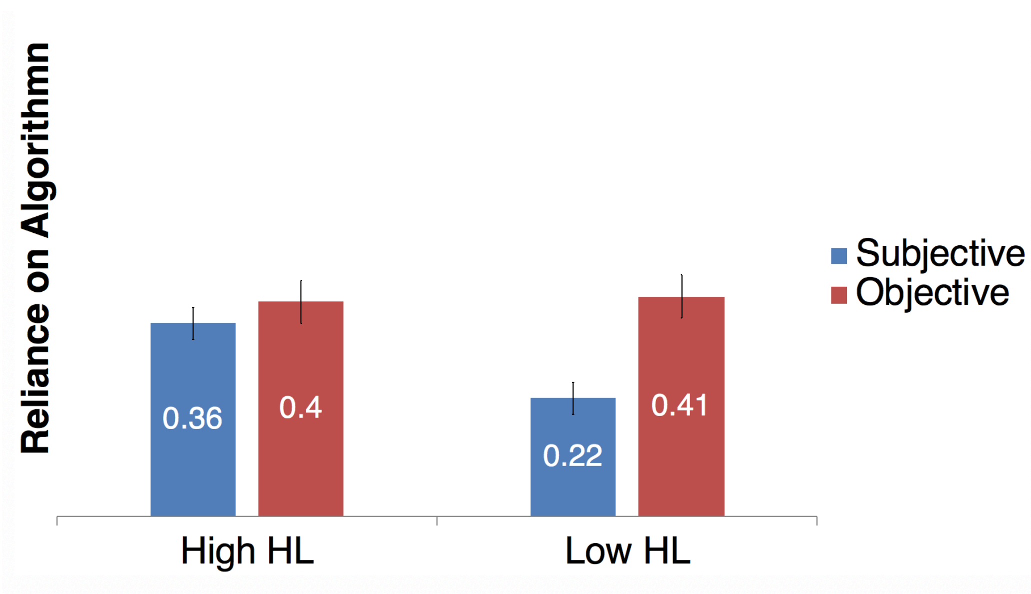


Figure 4. Task objectiveness increases reliance on algorithms when human-likeness is low but not high. Note: HL = human-likeness. Error bars represent standard errors.

Furthermore, I conducted another 2x2 ANOVA with belief in the algorithm's usefulness as the dependent variable. This revealed no main effect of task framing, $F(1,363) = .94$, $p = .339$, a marginally significant main effect of the algorithms' human-likeness, $F(1,363) = 2.75$, $p = .099$, and a significant interaction between these two factors, $F(1,363) = 4.10$, $p = .044$. The interaction pattern was the same as the previous interaction: when participants were told that algorithms have low human-likeness, the effect of task framing on perceived usefulness of the algorithm was significant ($M_{subjective} = 73.1$, $M_{objective} = 79.1$, $t(364) = 2.10$, $p = .037$). However, when participants were told that algorithms have high human-likeness, the effect of task framing was not significant ($M_{subjective} = 80.9$, $M_{objective} = 78.5$, $t(364) = .91$, $p = .366$).

Looked at another way, the effect of algorithm human-likeness on perceived usefulness was significant in the subjective condition ($M_{high\ HL} = 80.9$, $M_{low\ HL} = 73.1$, $t(364) = 2.54$, $p = .014$), but not in the objective condition ($M_{high\ HL} = 78.5$, $M_{low\ HL} = 79.1$, $t(364) = .29$, $p = .834$). This indicates that increasing algorithm human-likeness increases the perceived usefulness of algorithms for exactly the kind of task for which algorithms are typically not trusted or used.

The same ANOVA with discomfort as the dependent variable revealed no main effects but a significant interaction, $F(1,363) = 4.06$, $p = .045$. In the low human-likeness condition, discomfort with algorithms was higher in the subjective task condition ($M_{subjective} = 28.8$, $M_{objective} = 21.0$, $t(364) = 2.02$, $p = .044$). In the high human-likeness condition, discomfort was roughly equivalent across both task framings ($M_{subjective} = 26.2$, $M_{objective} = 30.3$, $t(364) = .91$, $p = .366$). Furthermore, the belief that algorithms make humans less distinct from machines was positively associated with discomfort ($\beta = .42$, $p < .001$) but was not associated with the perceived usefulness of the algorithm ($\beta = .04$, $p = .321$). Finally, as in Study 4, discomfort on its own was negatively associated with reliance on the algorithm ($\beta = -.001$, $p = .040$), while perceived usefulness on its own was positively associated with reliance ($\beta = .003$, $p < .001$). However, discomfort became non-significant when controlling for usefulness ($\beta = -.0006$, $p = .486$), while usefulness remained significant ($\beta = .003$, $p = .002$). This result suggests that consumers' discomfort with algorithms, stemming partly from decreasing human distinctiveness, has effects on consumers' willingness to use algorithms *ceteris paribus,* but that the effects of this discomfort are diminished when the algorithm is perceived as being highly useful.

These results also provide further evidence that algorithms are relied on less for tasks that seem subjective and suggest an additional method of eliminating this effect by increasing awareness of algorithms' affective human-likeness in terms of abilities normally seen as distinguishing humans from machines. Importantly, these results also help to tease apart the competing hypotheses regarding the role of algorithms' human-likeness in shaping consumers use of algorithms. Whereas social identity theory would suggest that algorithms that challenge the distinctives of humans from machines would create negative evaluations of those algorithms, a more cognitive perspective based on consumers' beliefs about algorithms' usefulness suggests that decreasing human-machine distinctiveness would increase consumers' use of algorithms by making algorithms seem more useful. The fact that I found support for the second of these hypotheses suggests that usefulness beliefs are stronger determinants of reliance on algorithms than discomfort stemming from intergroup challenges.

GENERAL DISCUSSION

As algorithms become increasingly capable of outperforming humans at tasks ranging from making recommendations (e.g. for music, movies, and stocks) to diagnosing diseases and driving cars, a key issue is whether (or at least when or how quickly) and for what purposes humans will trust and use them. This Essay explored several aspects of this question. Specifically, in a series of 6 experiments with over 56,000 participants in total, I have studied how trust in and use of algorithms varies depending on how both the task at hand and how algorithms are perceived.

Study 1 found that trust in algorithms for a given task is negatively related to perceived subjectivity. Study 2 replicated these findings in a field study, finding that consumers click on ads for algorithm-based advice less than on ads for human-based advice when the task is subjective (dating advice), but not when the task is objective (financial advice). Study 3 tested the effect of making consumers explicitly aware of an algorithm's superior performance compared to humans, finding that such awareness is indeed a powerful influence on willingness to use algorithms. However, awareness of superior performance is not sufficient for creating a true preference for algorithms over humans – only indifference between the two – and is particularly ineffective for tasks that are more subjective. Study 4 manipulated perceived task objectiveness, finding that re-framing subjective tasks as being amenable to quantification and measurement increases trust in algorithms for those tasks. Study 5 replicated this finding in a field study. Finally, Study 6 showed that actual reliance on algorithms in an incentivized task is also lower when the task is seen as subjective, but that this effect can be eliminated by providing real examples of algorithms with affective abilities.

Limitations and Directions for Future Research

I identify several limitations of this work. First, half of the studies relied on participants' reports of what they intended to do rather than direct evidence of what they did. More studies like Studies 2, 5, and 6 which used real behaviors as dependent variables are needed to further calibrate real-world reactions to algorithms. Related to this, attempts to assess the costs of choosing not to rely on algorithms whose performance is superior to humans may be a valuable pursuit for future research. That being said, it is encouraging that the self-report findings were replicated in field settings. Importantly, potential concerns about the results being driven by demand effects can be eliminated by the use of both field studies and incentive-compatible behaviors to replicate the effects observed in the self-report studies.

The manipulation of human-likeness in Study 6 also leaves open the possibility of alternative manipulations of this construct in future research. I chose to manipulate specifically the affective dimension of human-likeness, because this is the most closely relevant dimension to the performance of subjective tasks. Indeed, I expected (and found) that increasing affective human-likeness would make algorithms seem more effective and useful for subjective tasks. In this sense, affective human-likeness is inextricably linked to the usefulness of the algorithm for subjective tasks, since performing such tasks requires affective abilities. While I manipulated the perceived effectiveness of the algorithm for the tasks in question directly in Study 3, the manipulation of human-likeness in Study 6 is not specific to the task in question (i.e., forecasting stock prices), but was more domain-general. This had the intended effect of making the algorithm seem more useful for specific subjective tasks. However, this manipulation also had the expected effect of increasing discomfort with the algorithm. The purpose of Study 6 was therefore to pit the effects of usefulness and discomfort against each other, since both are affected by human-likeness and both were expected to impact reliance on algorithms, but in opposite directions. Nevertheless, alternative operationalizations of affective human-likeness would be worth testing as well, as would increasing algorithms' cognitive human-likeness, or even their physical human-likeness in order to test how these other dimensions might affect reliance on algorithms for different kinds of tasks.

Another limitation is that the descriptions of the algorithms were quite basic. More realistic presentations might include brochures, ads, websites, and videos. In addition, different presentation modes (in terms of both what and how information is presented) could be examined: use of algorithms may depend as much or more on how they are presented and accessed as on what is said about them. One aspect of this is the potential role of social influences on both individual decisions and overall adoption patterns (e.g., how do the p and q coefficients of the Bass diffusion model, which measure advertising and word of mouth effects on adoption, differ between algorithm adoption and adoption of other consumer and industrial products; Bass 1969).

Our finding that consumers are relatively averse to algorithms that are used for subjective tasks is particularly relevant in light of the current trend toward affective computing, which is a growing industry intent on creating explicitly emotional algorithms and building them into products from driverless cars to refrigerators to digital personal assistants (Goasduff 2017; Kodra et al. 2013). These results suggest that consumers will likely be skeptical about the emotional abilities of such algorithms, but that convincing demonstrations of their effectiveness may ultimately increase willingness to use them for tasks normally thought to be “incompatible” with algorithms or computers. Future research should explore how different ways of presenting emotional algorithms to consumers impacts their acceptance of those algorithms.

Future research can also explore additional factors that shape use of algorithms. For example, several factors that are not explicitly related to the nature of the algorithm or to the algorithm’s performance might affect use, such as concerns about privacy or simply the enjoyment of performing a task oneself. While it is beyond the scope of this Essay to explore all such factors, I conducted a small survey to begin exploring such beliefs. Specifically, I asked 80 MTurk participants to report how much they agreed with several potential concerns if they were to use an algorithm for each of 4 tasks, as well as whether they would prefer to use an algorithm or a human for the tasks (predicting personality based on Facebook likes, predicting joke funniness, planning a wedding, and predicting the price of stocks). For example, participants were asked whether using an algorithm for planning a wedding would involve privacy concerns, or whether the task requires emotion. Table 5 depicts the results of this survey. Interestingly, concerns about privacy implications and companies using algorithms for marketing purposes were not significant predictors for any task. In contrast, the enjoyment of doing the tasks oneself, feeling less control over the task, feeling bad about oneself if using an algorithm, the belief that the task is related to what it means to be human, and the belief that the task requires emotion were each significant predictors of preference for using an algorithm for at least two of these tasks. Future research is therefore needed to explore the role of these and other non-performance-related concerns in shaping consumers' use of algorithms.

	Preference for using humans for:			
	Personality prediction	**Joke funniness prediction**	**Wedding planning**	**Stock price prediction**
Privacy concerns	-.05	.05	.09	.22
Marketing concerns	.02	.05	.10	.27
Enjoy doing the task myself	.08	**.36**	**.47**	.27
Feed bad about myself	.22	**.36**	**.47**	.29
Feel less control	.28	**.44**	**.29**	**.34**
Requires emotion	**.35**	**.54**	**.48**	**.51**
Related to being human	.16	**.39**	**.65**	**.34**

Table 5. The effects of additional factors on preference for using humans relative to algorithms for different tasks. Note: standardized regression coefficients are displayed; significant effects are in bold). Positive coefficients indicate that greater concern with the factor in question is associated with greater preference for using a human (lower preference for using an algorithm).

Additional broader questions revolve around the potentially detrimental effects of increasing reliance on algorithms. For example, could an increased reliance on algorithms diminish people's capacity to think on their own and solve problems creatively, or to perform the tasks that have been outsourced to algorithms? Could it diminish the utility and satisfaction that people receive from accomplishing tasks on their own? Furthermore, some have argued that the increasing use of algorithms in society can entrench economic and social inequalities by building discrimination into inflexible models applied on a large scale in contexts such as parole, hiring, and credit decisions (O'Neil 2016). Any attempts to increase the use of algorithms in order to improve outcomes for consumers and society should be mindful of these concerns, striving to ensure that the promoted algorithms are both effective and fair.

Conclusion

I examined consumers' willingness to use algorithms to perform tasks in several areas and examined ways to increase their use. This Essay highlights the important role of perceived task objectiveness in shaping consumers' trust in and use of algorithms. While there appears to be an inexorable trend toward increasing use of algorithms, the pace at which they are adopted – as well as the areas where they will be adopted first – depends on a number of interrelated factors including in which areas companies develop algorithms for general use, how they market them, and how soon customers trust in and become comfortable with the idea of using algorithms to outsource decisions that affect their lives, in ways large and small. This research suggests that marketers face a challenge in balancing between increasing the capabilities of algorithmic products and services into subjective domains while also addressing consumers' lay beliefs that algorithms are ineffective at such tasks. These results provide several practical strategies for achieving this balance.

CHAPTER 4: REACTIONS TO ROBOTS

Robots – intelligent, physically embodied machines that can sense and manipulate their environment and perform tasks autonomously – are becoming increasingly prevalent in many domains of business and consumer behavior (Simon 2018). Today, there are already more than 1.5 million robots used worldwide in manufacturing alone. However, recently the market for social robots (which are intended to interact directly with consumers at home and in retail and service contexts) has been growing seven times faster than the market for manufacturing robots (Business Insider 2015), reaching $5.4 billion in sales by the end of 2017 and expected to triple to $14.9 billion within the next five years (Business Wire 2017). Consumers can already purchase social robots that perform chores, monitor young and elderly people, engage in conversations, and act as companions and assistants (Gibbs 2016a). Outside the home, social robots are being used in a wide variety of contexts including retail stores, restaurants, hotels, and hospitals (Dass 2017; Nguyen 2016; Simon 2015).

Social robots vary significantly in terms of how human-like they look. Some social robots, such as Jibo (sold for use in the home and recently featured on the cover of *Time* magazine as one of the best inventions of 2017) allude to human features by using a large round screen for a face but lack any recognizably human body. Other robots such as Pepper (a robot used in retail stores, hotels, restaurants, and airports) are more humanoid in their appearance, having a more human-like face as well as extremities resembling human legs, arms, and hands. Finally, some robots such as Erica (recently employed as a hotel concierge and television news anchor in Japan) are designed to be exact replicas of humans and are increasingly difficult to distinguish from real humans.

In addition to increasing physical human-likeness, social robots are also becoming more human-like in terms of their mental abilities. Progress in AI has already endowed computers (and therefore robots, which are physically embodied computers) with the ability to understand and produce hundreds of human languages and engage in complex conversations (Joshi 1991; Perera and Nand 2017), to drive cars and diagnose diseases (Simonite 2014), and even to understand human emotion by analyzing facial expressions and tone of voice (Kodra et al. 2013). Experts estimate that AI will have a 50% chance of achieving human-level intelligence within the next 20-25 years (Müller and Bostrom 2016). Robots are thus becoming more human-like both in terms of their bodies and their minds.

As increasingly human-like robots are entering the consumer marketplace, it is imperative to understand how consumers will react to such robots and ultimately how robots can provide value for consumers and firms. There already exists a small literature exploring how robots' physical appearance impacts consumers' affective reactions to robots, but the question of how robots' mental abilities impact consumers' and firms' reactions to robots remains almost entirely unexplored. I contribute to answering this question with five studies. I find that increasing the perception that social robots with human-like appearances also have human-like minds makes them seem more useful and increases consumers' comfort with them, willingness to patronize businesses that employ them, and interest in promoting their development. In effect, perceiving minds in human-like robots makes them more valuable to both consumers and firms. This increased value is partly due to the fact that minds enable empathy, such that robots with minds can better understand what humans are thinking and feeling.

Our research contributes to the literature on product anthropomorphism by showing how both a product's physical appearance (the focus of existing research in this area) and its mental abilities (a novel factor) interact in shaping consumers' reactions to the product. This interaction between "body" and "mind" also contributes to the literature on mind perception, which thus far has not incorporated physical appearance as a factor in the process of mind perception. I contribute to the literature on human-robot interaction by providing one of the first empirical tests of human reactions to robots with very highly human-like appearances, as these robots are only just being developed and reaching the marketplace now. Finally, I contribute to marketing practice by demonstrating how to increase social robots' perceived mind, thereby increasing the value that social robots can provide to both consumers and firms.

ROBOT HUMAN LIKENESS

Prior research on the relationship between robots' human-likeness and human evaluation of robots has focused on physical human-likeness, guided by the "uncanny valley hypothesis." Nearly 50 years ago, Masahiro Mori, a Japanese roboticist, wrote an influential paper speculating that making robots look more human-like is beneficial only up to a point, after which they become *too* human-like and elicit strongly negative reactions (Mori 1970). The "valley" thus refers to the worsening of reactions as robots move from moderately human-like to very (but not perfectly) human-like, and the subsequent improving of responses as human-likeness approaches perfection (i.e., actual humans). Mori therefore advised that social robots should be designed to have a moderate degree of human likeness in order to avoid the negative reactions elicited by very human-like robots.

In the years since Mori's paper, research has produced inconsistent findings regarding whether the uncanny valley exists in the form originally proposed. For example, while some studies have found support for a non-linear relationship between a robot's human-likeness and affective reactions to the robot (Ferrari et al. 2016; Mathur and Reichling 2016), several others have found no such pattern (Bartneck et al. 2009; Rosenthal-Von Der Pütten and Krämer 2014; Zlotowski, Proudfoot, and Bartneck 2013). Indeed, two recent reviews of research on the uncanny valley concluded that "although the notion of the uncanny valley is plausible and is supported by plentiful anecdotal evidence, rigorous controlled studies have yielded mixed support for its existence" (Wang et al. 2015) and "it is surprising that empirical evidence for the uncanny valley hypothesis is still ambiguous if not non-existent" (Kätsyri et al. 2015).

These inconsistent findings can likely be explained by the fact that Mori's original hypothesis was presented as a broad, somewhat vague idea without precisely defined constructs and without any data. This inevitably resulted in subsequent research operationalizing the two key variables – the robots' human-likeness and human reactions to the robots – inconsistently. For example, the original hypothesis operationalized "reactions" in terms of feelings of eeriness and creepiness (Mori 1970), while more recent research has measured feelings of unease (Gray and Wegner 2012), likeability of the robot (Mathur and Reichling 2016), and repulsion (Ferrari et al. 2016). Despite these inconsistencies, however, one aspect of the uncanny valley hypothesis that has been reliably demonstrated is that highly, but imperfectly human-like robots elicit negative affective reactions (Ferrari, Paladino, and Jetten 2016; Mathur and Reichling 2016).

Two classes of explanations have been proposed for negative reactions to human-like robots: perceptual and cognitive explanations (Wang et al. 2015). Perceptual explanations focus on how aesthetic imperfections in robotic faces or bodies create negative reactions. For example, the *evolutionary aesthetics hypothesis* explains negative reactions to human-like robots by suggesting that evolution has shaped human preference for physical appearances that signal health and fitness, and that humanoid robots lack such an appearance because of their aesthetic imperfections (Hanson 2005). Thus, any physical imperfection that signals potential ill health or increases the salience of disease or death can produce feelings of unease and uncanniness. Similarly, the *pathogen avoidance hypothesis* proposes that physically imperfect faces are indicative of transmissible diseases (MacDorman et al. 2009), and the *mortality salience hypothesis* suggests that such faces remind viewers of their own mortality (Ho et al. 2008) both of which then create negative reactions.

Cognitive explanations include the *violation of expectations hypothesis*, which was the explanation proposed by Mori in the first Uncanny Valley paper, and which argues that human-like robots create an initial expectation of a human but then fails to meet those expectations (Mori 1970). More recent research has found that mismatches between physical appearance and motion (Saygin et al. 2012) and between face and voice (Mitchell et al. 2011) both produce negative reactions towards human-like robots. In other words, if a robot looks very human-like but does not move like a human or has a human-like face but a robotic voice, the robot has created an expectation of humanness but then failed to meet that expectation due to other, non-human-like features. A similar explanation is the *category uncertainty hypothesis*, or the idea that negative reactions are caused by a difficulty in categorizing something as a robot or a human (Yamada, Kawabe, and Ihaya 2013).

As robots continue to become more human-like, and indeed approach perfect human-likeness in the physical sense, many of the aesthetic causes of negative affective reactions cited above will disappear, as robots become aesthetically indistinguishable from humans. However, I suggest that a focus on physical human-likeness alone is insufficient for understanding and improving consumers' reactions to social robots. There are many reasons why even robots with perfect physical human-likeness could still elicit negative reactions: category uncertainty would be perhaps even greater since it would be harder to distinguish between robots and humans (Yamada et al. 2013); such robots could be particularly threatening to human jobs, safety, and distinctiveness (Ferrari et al. 2016), and could create strong expectations of humanness that would then be violated if the robot is not perfectly human-like in non-aesthetic ways. This leads to our first hypothesis:

H1: Robots with perfect physical human-likeness will elicit more positive affective reactions than robots with high but imperfect physical human-likeness, but more negative affective reactions than actual humans.

Insofar as even robots with perfect physical human-likeness still elicit relatively negative reactions, non-aesthetic factors will become important in further improving consumers' reactions to highly human-like robots. In this Essay, I focus on mind perception as one such factor. We know that highly (and especially perfectly) human-like robots can create the initial expectation of a human and that failure to meet those expectations creates negative reactions (Mitchell et al. 2011; Saygin et al. 2012). Meeting the expectation of being human, however, requires not only a body, but also a mind.

MIND PERCEPTION

Consumers perceive minds along two primary dimensions: agency (the ability to plan and act autonomously) and experience (the capacity to feel emotions and sensations; Gray, Gray, and Wegner 2007; Waytz et al. 2010). Robots are perceived as having moderate levels of agency but virtually no capacity for experience (Gray, Gray, and Wegner 2007). These dimensions of mind correspond to real abilities that modern robots do in fact possess. In terms of agency, robots can increasingly engage in complex conversations, behaviors, and decisions without being controlled or supervised by a human (Beer, Fisk, and Rogers 2014; Vázquez et al. 2017). Regarding experience, robots can accurately detect and classify human emotions by analyzing facial expressions and tone of voice, and use that data in order to express their own "emotional reactions," by tailoring their responses to the emotional state of the human they are interacting with: making jokes, expressing sympathy, and so on (Khatchadourian 2015; McDuff et al. 2016; Picard 2011).

These abilities map closely (albeit imperfectly) on to agency and experience as they are defined in the psychology literature. I will show that robots that display these capabilities (or are described as having these abilities) are perceived has having greater autonomy and emotional experience, respectively. Furthermore, while there are deep philosophical questions about whether robots can truly have autonomy or emotional experiences (Dennett 1997), I do not take a stand on this debate. Instead, I exploit the reasonable arguments to be made on both sides of the debate to create our manipulations of perceived minds in robots. These arguments will be explained in the methods section of Study 2.

How would the perception that robots have human-like minds impact consumers' reactions to robots? That is the central research question of this Essay. I first define what I mean by reactions. In line with the model proposed in Essay 1, I measure two key dimensions of reactions: the perceived usefulness of robots, and consumers' comfort with robots. Both usefulness and comfort contribute to the value that social robots can provide to consumers and firms. Virtually all models of technology adoption emphasize perceived usefulness as one of the strongest determinants of adoption (Venkatesh et al. 2003). Understanding what shapes perceived usefulness is therefore a fundamentally important goal for the developers and marketers of technologies. I measure comfort as well because research on consumer perceptions of human-like robots has typically focused on measuring affective reactions such as creepiness or discomfort (Kätsyri et al. 2015; Wang et al. 2015). In order to build on this literature, I therefore also measure affective reactions, focusing specifically on comfort with robots both in general and in specific consumption contexts. I propose that mind perception will increase both consumers' comfort with social robots as well as their perceived usefulness.

Comfort. Building on the violation of expectations hypothesis, I suggest that robots with highly human-like physical appearances will prime expectations of humanness, an important component of which is a human-like mind. Something that looks like a human, in other words, may also be expected to have a mind like a human; to think and feel like a human. This basic congruence between physical human-likeness and mental human-likeness should increase comfort. Research has indeed shown that congruence between a product's features and the category schema of which it is a part creates positive affect which is then transferred to the product itself (Meyers-Levy and Tybout 2002). More recent work has confirmed this finding specifically in the context of product anthropomorphism, showing that anthropomorphizing a product by having it portray itself in the first person primes the category schema of "human," which leads to more positive product evaluations when the product seems to be smiling (since smiles were found to be congruent with the human schema; Aggarwal and McGill 2007). I extend this logic by suggesting that a robot's perceived mind also contributes to its congruence with the human schema, ultimately increasing comfort with the robot. Robots with both dimensions of mind are by definition more congruent with the human schema than robots with a single dimension and should therefore produce the greatest comfort, although robots with either dimension are more congruent than robots with neither and should therefore produce greater comfort accordingly. Our main focus, however, is comparing robots with both dimensions of mind to robots with neither.

H2a: Perceiving a physically human-like robot as having a human-like mind (vs. no mind) will increase consumers' comfort with such robots.

Increasing a robot's perceived mind may also boost comfort via a second path. Entities with minds are capable of *empathy*, specifically the two components of empathy known as perspective-taking and sympathy. Perspective-taking refers to the capacity to understand other peoples' beliefs, intentions and thoughts, while sympathy (also called compassion in the empathy literature) refers to a feeling of concern for another person's suffering accompanied by the motivation to help (Singer and Klimecki 2014). Perceiving a robot as having a human-like mind, capable of acting with autonomy and of understanding, expressing, and even experiencing emotions, should increase the robot's perceived ability to engage in both perspective-taking and sympathy for humans. These abilities should in turn increase consumers' comfort with robots.

H2b: A robot's perceived empathy for humans will mediate the effect of mind perception on comfort with robots.

Usefulness. I expect that perceiving social robots as having human-like minds will also increase their perceived usefulness. Intuitively, having greater autonomy makes robots more useful almost by definition, since they can perform more tasks without direct supervision or control by humans. Being able to understand and express emotion also enables a robot to perform more useful tasks. I expect that empathy plays a key role in the effect of mind perception on usefulness as well. Social interactions are facilitated by empathy: interacting with someone or something that can take one's perspective and feel for one's suffering fosters social bonds, facilitates group coordination, and increases prosocial behavior (Galinsky, Ku, and Wang

2005; Oswald 1996; Rosenthal-von der Pütten et al. 2014). Social robots, whose fundamental purpose is to engage in effective social interactions with humans, should therefore be seen as (and indeed *be*) more useful when they are seen as having a human-like mind, and this effect should be explained partly by perceived empathy.

H3a: Perceiving a robot as having a human-like mind (vs. no mind) will increase the robot's perceived usefulness.

H3b: A robot's perceived empathy for humans will mediate the effect of mind perception on usefulness.

Note that H2a refers specifically to physically human-like robots while H3a does not. This is because I expect that the effect of mind perception on comfort – but not on usefulness – will be moderated by physical human-likeness. Having a mind should make robots seem more useful regardless of what they look like. In terms of comfort, robots with low or moderate physical human-likeness should not elicit the category schema of human, and the presence of a mind in such robots should therefore not increase schema congruity. Indeed, research has found that consumers are *less* comfortable when robots low in physical human-likeness are described as having a mind (specifically the experience dimension of mind; Gray and Wegner 2012). In contrast, robots with high physical human-likeness are more likely to elicit the category schema of human, such that having a human-like mind is congruent with the elicited schema.

H4: The positive effect of mind perception on comfort with highly human-like robots will be reduced for low human-likeness robots.

Finally, I also expect that these effects will have practical marketing implications. As mentioned, a technology's perceived usefulness is one of the strongest factors in shaping the technology's adoption (Venkatesh et al. 2003), and the well-established discomfort that consumers feel with human-like robots is an obvious barrier to the effective use of such robots in consumption settings. Increasing perceived usefulness and comfort via mind perception will therefore increase the value that consumers and firms can obtain from social robots. I will measure these downstream consequences in various ways throughout the Essay.

OVERVIEW OF THE STUDIES

Study 1 shows that increasing physical human-likeness from high to perfect does indeed improve consumers' affective reactions to social robots, but that reactions remain far more negative than reactions to actual humans. The remaining studies focus on how mind perception can improve reactions to highly human-like robots. Study 2 provides causal evidence for the benefits of mind perception, showing that perceiving robots as having human-like minds increases consumers' comfort with the use of robots in stores and restaurants. This study also measures two downstream consequences of increased comfort: willingness to donate to an organization that promotes the development of robots, and evaluations of companies that employ robots. Study 3 shows that both dimensions of mind – agency and experience – have positive effects on consumers' comfort with and perceived usefulness of social robots. Study 4 shows that the effect of mind perception on comfort – but not on usefulness – is moderated by the robot's physical appearance. This study also shows that the robot's perceived empathy for humans can partially account for these effects. Study 5 shows that perceiving minds in human-like robots increases comfort on a physiological level.

STUDY 1

The bulk of this Essay focuses on the effects of mind perception on consumer reactions to robots. This focus is motivated by the idea that highly human-like robots elicit negative affective reactions, and that even perfectly human-like robots (in a physical sense) will still elicit more negative reactions than actual humans, suggesting that non-aesthetic factors such as mind perception are also important for understanding and improving consumer reactions to robots. This study provides empirical support for that idea, thus helping to motivate the remaining studies.

Method

Participants and design. I recruited 800 American participants (51% female, mean age = 33) from Prolific Academic, a crowdsourcing website where participants are less familiar with common experimental paradigms and more honest than participants on Mechanical Turk (Peer et al. 2017). Participants were randomly assigned to rate several robots either on human-likeness or on their affective reactions to the robots.

Procedure. Participants were shown images of robots compiled by the Anthropomorphic Robot Database (Phillips et al. 2018). I chose 25 of these robots varying in human-likeness from very low to very high. 200 of the participants rated the robots on how human-like they looked overall. This allowed us to divide the 25 robots into five quintiles based on their overall human-likeness. The remaining 600 participants were shown five of the robots (one from each of the five quintiles) or one of four human beings. Participants were either told that these humans were humans or were in fact advanced humanoid robots. This allows us to compare evaluations of perfectly human-like robots to evaluations of actual humans. Please see Appendix B for all stimuli used in this Essay.

As our measure of affective reactions, I asked participants how creepy each robot seemed and how comfortable they would be interacting with each robot. Both items used 0–10 scales anchored at “not at all” and “completely.” I excluded 99 participants (16.5%) who reported suspicion that the humans portrayed as robots were not in fact robots, although the pattern of results in unchanged if I include these participants.

Results and Discussion

In order to test for a U-shaped relationship between human-likeness and consumer reactions, I conducted a two-lines test, which is a more accurate and valid method of testing for such relationships than testing for a quadratic term in a regression (Simonsohn 2018). This test estimates separate regression lines for low and high values of the x-variable (using an algorithm to determine the boundary point between low and high that maximizes the overall statistical power of the test) and is significant if the two lines are individually significant and opposite in sign. The results of this test are depicted in Figure 5 and show that increasing human-likeness leads to more negative reactions until the point 8.14 (slightly lower than the human-likeness of the most human-like robots that currently exist; $\beta = -.52$, $p < .001$). After this point, further increases in human-likeness lead to improved reactions, $\beta = 1.89$, $p = < .001$). Note that this analysis did not include the humans portrayed as humans.

The perfectly human-like robots in this study were evaluated more negatively than actual humans ($M_{Human} = 8.70$ vs. $M_{PerfectRobot} = 6.87$, $t(503) = 9.65$, $p < .001$), but more positively than the robots in the fifth (highest) quartile of human likeness (M_{Q5} = 4.16, $t(553) = 11.64$, $p < .001$), as well as the robots in the fourth quartile (M_{Q4} = 4.84, $t(416) = 7.39$, $p < .001$), third quartile (M_{Q3} = 6.17, $t(416) = 2.72$, $p = .007$), and second quartile (M_{Q2} = 6.05, $t(416) = 3.07$, $p = .002$), but more negatively than the robots in the first (lowest) quartile of human likeness (M_{Q5} = 7.48, $t(416) = -2.48$, $p = .013$).

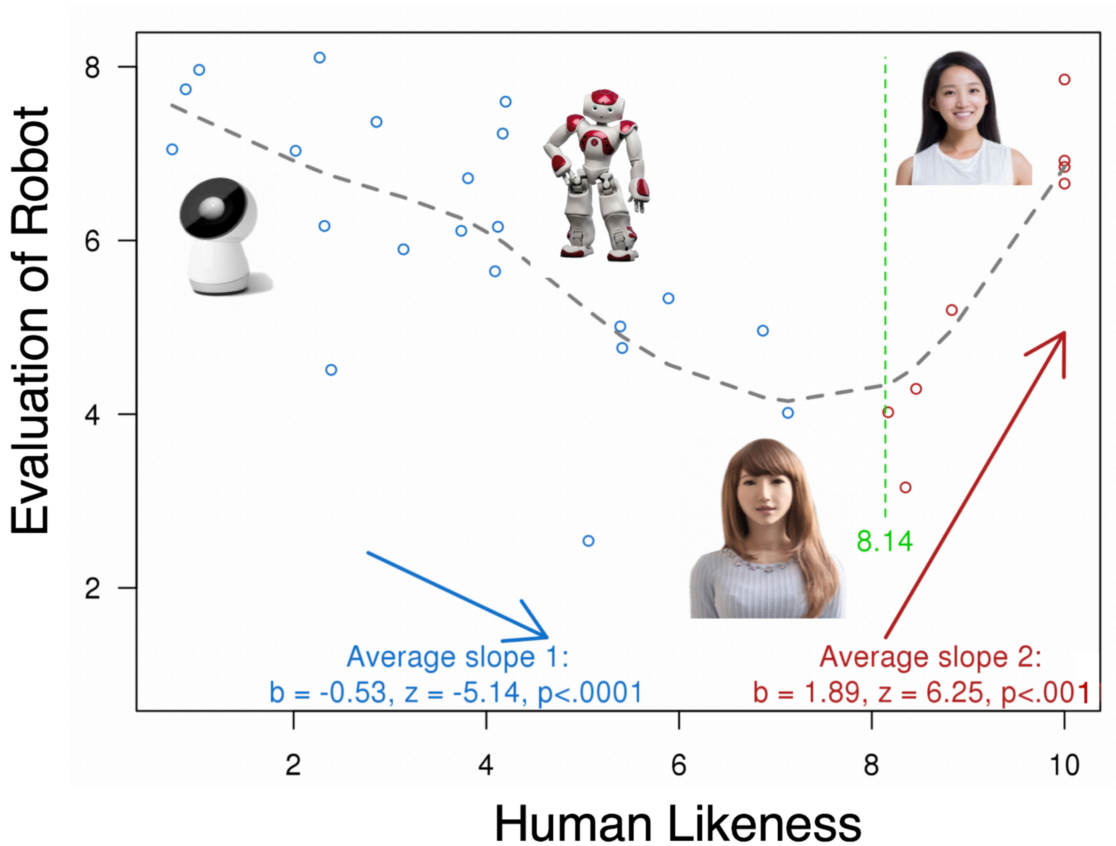


Figure 5. Human-likeness first elicits more negative reactions, then more positive reactions. Note: Individual points on the graph represent the individual robots used as stimuli.

It is worth noting that I do not observe the initial improvement in evaluations as human-likeness moves from low to moderate that Mori predicted when he proposed the uncanny valley hypothesis. The reason for this may be that Mori included industrial robots in his discussion, which have extremely low human-likeness, whereas our stimuli included only social robots that have at least some subtle allusions to human-likeness, such as a large round screen that alludes to a face. The initial improvement that Mori predicted may therefore only occur when non-human-like industrial robots are included as stimuli.

Existing research on the uncanny valley phenomenon has demonstrated that reactions to robots become more negative as human-likeness increases from low to high (Ferrari et al. 2016; Mathur and Reichling 2016). The results of study 1 replicate this finding and make a novel contribution by showing for the first time what happens to reactions when human-likeness approaches perfection.

This pattern suggests that firms who choose to employ social robots may be better off employing robots with low-to-moderate human-likeness. However, some firms are clearly pushing towards robots with ever greater human-likeness (i.e., Hanson Robotics and Sanctuary AI). The remaining studies therefore focus on testing the hypothesis that increasing the perceived minds of such robots will improve consumers' reactions to them, helping to bring them out of the uncanny valley.

STUDY 2

This study tests H2, that perceiving robots as having human-like minds will increase consumers' comfort with robots, relative to perceiving them as lacking human-like minds. I also measure two downstream consequences that should be related to comfort: willingness to donate to a pro-robot organization, and evaluations of companies that employ robots.

Method

Participants and design. 100 American MTurk participants (46% female, mean age = 36) were assigned to watch one of two videos.

Procedure. The videos featured a professor explaining that robots either could or could not have a mind in the same way that humans do. The arguments were taken from the philosophical literature on mind and consciousness, specifically on the "dualism vs. physicalism" debate that portrays the human mind as having non-material aspects that cannot be replicated in a machine ("dualism") or as something explainable entirely in terms of physical brain processes that can be replicated in a machine ("physicalism"; Dennett 1994; Searle 1995). This manipulation therefore represents a high-level operationalization of mind perception, or the general belief that robots can in principle have minds like humans do. I will therefore refer to these conditions as the "mind" and "no mind" conditions. Subsequent studies will use more concrete manipulations of mind perception that describe the specific capacities that make up a mind.

After watching one of the videos, participants were asked to spend two minutes writing about why the arguments in the video were likely to be true, which is a technique known as the "saying-is-believing technique" commonly used in psychology research to increase engagement with stimuli and facilitate attitude change (Higgins and Rholes 1978; Aronson, Fried, and Good 2002; Yeager et al. 2016). I excluded 4 participants who failed to write a single comprehensible sentence. As a manipulation check, I asked participants whether they believed that robots could have a mind, on a 0 (not at all) to 10 (completely) scale.

All participants were then shown a picture of Erica, a social robot with high physical human-likeness (an overall human-likeness rating of 89.6 in the Anthropomorphic Robot Database). The picture was accompanied by an explanation that robots are approaching perfect human-likeness and are being used as employees in stores, restaurants, and hotels. I asked participants how comfortable they would be (a) shopping in a store and (b) dining in a restaurant where this kind of robot was employed, on 0–10 scales anchored at "not at all" and "completely."

Consumers who are more comfortable with the use of robots should also view companies employing robots more favorably and be more willing to promote their development. I therefore also asked participants how they would evaluate a company that employed this kind of robot (on 1–7 scales anchored at negative/positive, dislike/like, and bad/good). I then told participants that I would be donating $1 on behalf of each participant to an organization working on human-robot relations and asked them to decide which organization I would donate to on their behalf. They were given a choice between the American Society for the Prevention of Cruelty to Robots, which was described as working to advance the development of human-like robots, and the Center for the Study of Existential Risk, which was described as working to prevent the development of human-like robots. Both organizations are real.

In order to ensure that our two videos were equally convincing and believable, I also asked how engaging and convincing the video was and how knowledgeable the speaker seemed. All of these measures used 0 (not at all) to 10 (completely) scales. I finally measured participants' age, gender, and level of education.

Results and Discussion

The two videos seemed equally engaging and convincing, and the speaker seemed equally knowledgeable in both videos (see Table 6 for full results). The manipulation check showed that the videos successfully altered the belief that robots could have a mind.

I then tested whether the mind perception manipulation impacted comfort with robots in stores and restaurants, using a 2 x 2 mixed ANOVA with the mind perception condition and context (store vs. restaurant) as the independent variables. Mind perception had a significant effect on comfort ($F(1,94) = 9.45$, $p = .002$); context did not ($F(1,94) = .23$, $p = .630$). The interaction was not significant ($F(1,94) = .09$, $p = .768$). Consumers expected to be more comfortable with the use of robots in the mind condition (M = 6.63) than in the no mind condition (M = 5.35, $t(94) = 3.09$, $p = .002$). Reflecting this increased comfort, participants in the mind condition were more likely to choose the pro-robot organization for their donation and evaluated companies employing robots more positively (see Table 6). Comfort predicted both donation choice ($\beta = -.07$, $p < .001$) and company evaluations ($\beta = .41$, $p < .001$).

Finally, I checked whether the effects of mind perception on comfort interacted with participants' age, gender, or education. Condition did interact with age ($F(1,94) = 5.81$ $p = .768$). A floodlight analysis (Spiller et al. 2013) revealed that condition increased comfort among participants under the age of 37.3 (70.5% of the total sample) but had no effect among participants older than that. No other main effects or interactions were significant, including 3- and 4-way interactions.

	No Mind	Mind	Statistical Test
Engaging	6.64	6.05	$t = 1.05$, $p = .297$
Convincing	6.85	6.72	$t = .23$, $p = .819$
Knowledgeable	7.83	7.19	$t = 1.31$, $p = .194$
Mind (MC)	1.77	6.31	$t = 7.69$, $p < .001$
Store Comfort	5.20	6.40	$t = 1.90$, $p = .061$
Restaurant Comfort	5.25	6.79	$t = 2.34$, $p = .018$
Company Evaluation	4.11	4.87	$t = 2.25$, $p = .027$
Pro-robot Donation	44.2%	64.7%	$\chi^2 = 3.19$, $p = .074$

Table 6. Results of Essay 3, Study 2. Note: all scales were 0–10 except donation choice, which was binary.

The results of Study 2 provide initial evidence that increasing the perception that social robots can have human-like minds make consumers more comfortable with such robots. However, this high-level manipulation of mind perception leaves open the question of what precisely is meant by “mind.” The next study will therefore manipulate mind perception in more concrete terms, using an operationalization taken directly from the psychological literature on mind perception. It will also measure the perceived usefulness of robots as an additional dependent variable (H3).

STUDY 3

The next study has two primary goals. The first goal is to provide a more concrete operationalization of mind perception, defining a mind specifically in terms of the two primary dimensions identified in the mind perception literature (agency and experience). I therefore manipulate consumers' perceptions of robots as having these two dimensions of mind. The second goal is to test H3, that mind perception will increase the perceived usefulness of social robots.

Method

Participants and design. 282 participants (51% female, mean age = 34) were recruited from Prolific Academic and assigned to one of four conditions: no mind, autonomy only, emotion only, or complete mind.

Procedure. All participants were shown a picture of a highly human-like robot (Erica), and were informed that these kinds of robots could either (a) experience emotions like humans do but could not make autonomous decisions; (b) could make autonomous decisions like humans do but not experience emotions; (c) had both capacities, or (d) had neither capacity. When the robots were said to have autonomy and/or experience emotion, I explained that this is possible because the experience of emotion and the capacity for autonomy is created by a pattern of electrical activity in the human brain, which can be replicated in machines to produce the same phenomena in robots. Participants were asked to summarize the information they read about in order to increase engagement with the stimuli.

Participants then completed manipulation checks (i.e., reported how much this kind of robot seems capable of experiencing emotion and making autonomous decisions).

Perceived usefulness of the robot was measured with two items: "this kind of robot seems competent," and "this kind of robot seems useful." Comfort with the robot was also measured with two items: "how comfortable would you feel as a patient in a hospital where this kind of robot is employed as a [*nurse* / *hospital administrator*]?" I asked specifically about these two jobs because they differ in the amount of social interaction with patients but are in the same hospital context. This allows me to test whether the effects of mind perception generalize to a task in which the amount of social interaction is generally quite low.

Results and Discussion

I first excluded 12 participants who did not write at least one full sentence in their summary of the stimuli. For each dependent variable, I conducted regressions with the no mind control set as the reference group and the autonomy only, emotion only, and complete mind conditions dummy coded as 0/1. I use dummy variable regression instead of ANOVA because I am mainly concerned with the effect of having a complete mind vs. no mind, and with comparing the two single dimensions of mind to no mind.

Manipulation checks. First, the manipulation checks revealed that our manipulations were effective: perceived autonomy was lowest in the no mind condition, not significantly higher in the emotion only condition, and significantly higher in the autonomy only and complete mind conditions (see Table 7 for all means and significance levels). Similarly, perceived emotion was lowest in the no mind condition, not significantly higher in the autonomy only condition, and significantly higher in the emotion only and complete mind conditions.

Usefulness. The robot's perceived competence and usefulness had high internal reliability ($\alpha = .71$) and so were averaged and analyzed together as a measure of perceived usefulness of the robot. The robot seemed more useful in each of the three mind conditions than in the no mind condition.

Comfort. The type of job (nurse vs. administrator) did not interact with any of the conditions in shaping comfort with robots (p's > .703), so I averaged these two items as our measure of participants' comfort with robots ($\alpha = .92$). Participants were least comfortable with mindless robots, non-significantly more comfortable with robots having either autonomy or emotion alone, and significantly more comfortable with robots having a complete mind. Despite the non-significant interactions between task and conditions, however, Table 7 shows that the effect of having a complete mind was slightly stronger when the robot was described as a nurse vs. administrator.

Study 3 thus provides further support for H1, demonstrating that consumers are more comfortable with robots having either dimension of mind compared to neither dimension, and are the most comfortable with robots having a complete mind. This study also provides initial support for H3 by showing that perceiving minds in robots increases their perceived usefulness. The next study will begin to explore the mediating role of empathy and will test whether the robot's physical human-likeness moderates the effect of mind perception.

	No Mind	Autonomy Only	Emotion Only	Complete Mind
Autonomy	4.63_a	5.72_b	4.85_a	5.21_b
Emotion	3.40_a	3.77_a	4.54_b	5.61_b
Usefulness	5.53_a	6.30_b	6.72_b	6.71_b
Nurse Comfort	2.45_a	2.87_a	2.59_a	3.49_b
Administrator Comfort	2.65_a	2.99_a	2.84_a	3.36_a
Average Comfort	2.55_a	2.93_a	2.71_a	3.42_b

Table 7. Results of Essay 3, Study 3. Note: In each row, means with subscript "a" are not significantly different from the no mind condition; means with subscript "b" are significantly different from the no mind condition. All measures used 0–10 scales.

STUDY 4

In Study 4, I test whether perceiving robots as having more empathy for humans can explain the effects of mind perception on comfort (H2b) and usefulness (H3b). I also test whether the effect of mind perception on comfort is moderated by physical human-likeness (H4). Finally, as another downstream consequence of comfort and usefulness, I measure participants' willingness to shop in a store where robots are employed.

Method

Participants and design. 300 Prolific participants (45% female, mean age = 32) were assigned to one condition in a 2 (complete mind vs. no mind) x 2 (low vs. high human-likeness) design.

Procedure. Mind perception was manipulated using the videos from Study 2 and by explaining that mind refers to autonomy and emotion specifically, defined as in Study 3. Participants summarized the stimuli and were then shown a picture of a robot with either moderate human-likeness (Pepper, rated as 42.2 out of 100 on overall human-likeness in the Anthropomorphic Robot Database; Phillips et al. 2018) or high human-likeness (Erica, rated as 89.6) and told that these robots are being used by businesses as receptionists, sales people, waiters, and more, and by individual people as social and romantic companions. A pre-test with 83 different Prolific participants confirmed that Erica was more effective at priming a human schema: the concept of "human" came to mind more easily when viewing the image of Erica (M = 4.35 on a 0–10 scale) than the image of Pepper (M = 3.06, $t(81) = 2.13$, $p = .037$).

I first measured participants' belief that the robot could have the two components of empathy ("how much do you think this kind of robot could understand what *you* are thinking and feeling," and "how much sympathy do you think this kind of robot would feel for you if you were suffering?"). All questions in this study used 0–10 scales.

Perceived usefulness of the robot was measured with the same two items used in the previous study (how competent and useful this kind of robot seems). I measured comfort in a more generalized or context-independent form than I did in Study 3, using a 3-item scale taken from prior research on the uncanny valley phenomenon, asking how much participants would feel uneasy, unnerved, and creeped out during an interaction with the robot (Gray and Wegner 2012).

Finally, I asked participants how willing they would be to go shopping in a store where the robot was employed as a sales clerk. This allows me to test how usefulness and comfort jointly shape willingness to interact with a robot in a consumption setting.

Results and Discussion

I first excluded 3 participants who failed to write a comprehensible sentence in their summary of the stimuli.

Robot's empathy for humans. Mind perception increased both the robots' perceived ability to take a human's perspective ($F(1,295) = 18.92$, $p < .001$) and to have sympathy for humans ($F(1,295) = 32.70$, $p < .001$). Neither human-likeness nor the interaction affected either form of empathy (F's < 2.44, p's $> .104$).

Usefulness. Usefulness was affected by mind perception ($F(1,295) = 10.99$, $p = .001$); the effect of human-likeness was not significant ($F(1,295) = .55$, $p = .461$), nor was the interaction ($F(1,295) = .36$, $p = .551$). Mind perception increased the perceived usefulness of the robots ($M_{mind} = 5.91$, $M_{no\_mind} = 5.08$, $t(295) = 3.32$, $p = .001$).

Comfort. Generalized comfort with the robots was marginally affected by mind perception ($F(1,295) = 3.12$, $p = .078$), significantly by human-likeness ($F(1,295) = 35.60$, $p < .001$), and significantly by the interaction term ($F(1,295) = 10.79$, $p = .001$). Having a mind decreased comfort with the low human-likeness robot ($M_{mind} = 5.28$, $M_{no\_mind} = 6.62$, $t(295) = 3.24$, $p = .001$) but directionally increased comfort with the high human-likeness robot ($M_{mind} = 4.38$, $M_{no\_mind} = 3.86$, $t(295) = 1.43$, $p = .156$).

Willingness to shop. Willingness to shop in a robot-staffed store was not affected by mind perception ($F(1,295) = .64$, $p = .423$) or by human-likeness ($F(1,295) = 1.76$, $p = .185$). However, there was a marginal interaction ($F(1,295) = 2.79$, $p = .096$). Having a mind did not affect intentions to shop with the low human-likeness robot ($M_{mind} = 6.09$, $M_{no\_mind} = 6.32$, $t(295)$

= .53, $p = .599$) but increased intended likelihood of shopping with the high human-likeness robot ($M_{mind} = 6.17$, $M_{no\_mind} = 5.35$, $t(295) = 1.81$, $p = .072$).

Mediation. Mediation analysis revealed that the perceived usefulness of the robot is more important than comfort both as an effect of mind perception and empathy and as a determinant of willingness to shop in a robot-staffed store. Comfort was not significantly affected by the robots' perceived empathy for humans ($\beta = .05$, $p = .429$) or by mind perception ($\beta = -.55$, $p = .094$; although the interaction with human-likeness was highly significant as reported above). In contrast, usefulness was affected by both the robots' perceived empathy ($\beta = .43$, $p < .001$) and by mind perception (($\beta = .85$, $p < .001$). Usefulness also had a stronger effect on willingness to shop ($\beta = .56$, $p < .001$) than comfort did ($\beta = .31$, $p < .001$). The effect of mind perception on usefulness was mediated by the robots' perceived empathy ($\beta = .62$, 95% CI = [.37, .92]). Furthermore, empathy and usefulness serially mediated the effect of mind perception on behavioral intentions ($\beta = .35$, 95% CI = [.20, .58]). The full model is displayed in Figure 6.

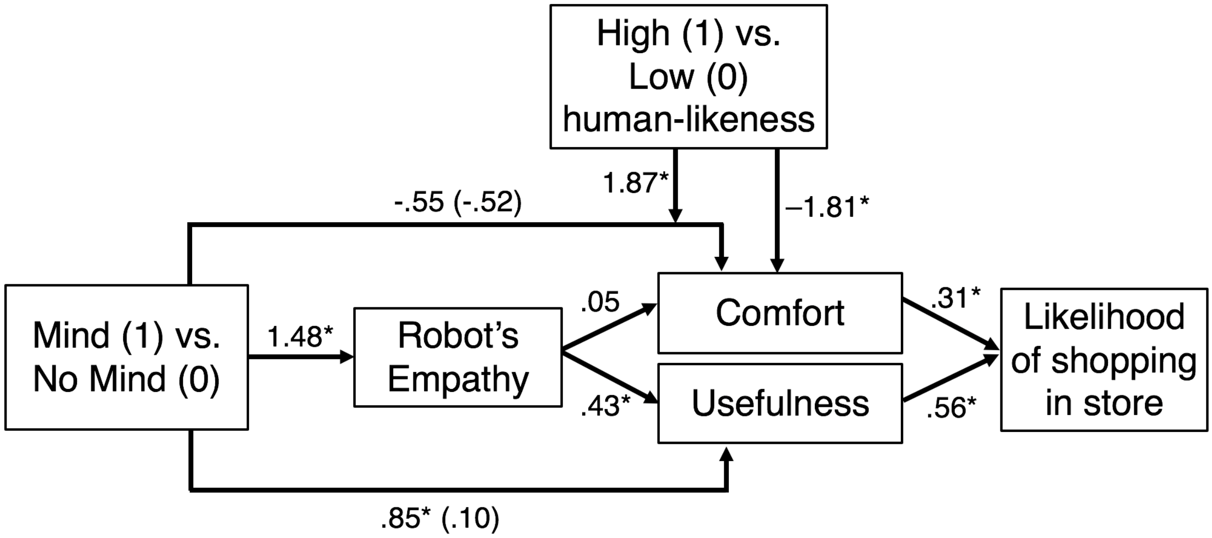


Figure 6. Model tested in Essay 3, Study 4.

Study 4 therefore confirms the value-increasing effects of mind perception in highly human-like robots and shows that these effects can be partially explained by the belief that robots with minds can have empathy for humans. It also shows that mind perception has uniformly positive effects on usefulness in particular for both high and low human-likeness robots but has opposite effects on comfort. This finding supports the role of congruity between physical and mental human-likeness in shaping consumer reactions to robots. It also illustrates a tension in the development and use of robots who should ideally be seen as both useful and as entities with whom consumers are comfortable.

STUDY 5

The results presented thus far are limited by the fact that they have relied mostly on self-reported attitudes and intentions, with the exception of participants' choice of donation in Study 2. I therefore conducted Study 5 in order to measure consumers' reactions to human-like robots using a more behavioral measure: physiological reactions (Morales, Amir, and Lee 2017).

Emotional or affective experiences consist of both psychological and physiological components (Barrett et al. 2006), and the physiological component of such experiences can be measured by electrodermal activity (EDA), or the electrical conductance of the skin. Specifically, EDA is defined as the change in the electrical properties of the skin in response to the secretion of sweat (Turpin and Grandfield 2010). EDA increases along with the self-reported emotional intensity of an experience (Lang et al. 1993; Manning and Melchiori 1974; Winton, Putnam, and Krauss 1984). Increased EDA has been linked specifically to heightened anxiety and arousal in response to negative emotional stimuli (Balconi, Falbo, and Conte 2012; Nikolić et al. 2018). This measure is therefore a good proxy for our concept of comfort as one important dimension of consumers' reactions to robots.

Method

Participants and design. 83 students at Columbia University (60.1% female, mean age = 24) participated in this study and were assigned to one of two conditions: no mind vs. complete mind.

Procedure. Upon arrival at the lab, a research assistant attached Biopac Ag-AgCL electrode sensors to the thumb and middle finger of participants' non-dominant hand in order to measure electrodermal activity. Participants were run one at a time. Following standard recommended procedure, I prepared the electrodes with isotonic gel to ensure reliable measurement (Lajante et al. 2012). The electrodes were connected to the Biopac MP150 data acquisition unit, which records the data collected from the electrodes and transmits it to a computer for storage and analysis.

EDA has two main components. The first is called tonic activity, which changes relatively slowly in response to affective stimuli over a period of 10 seconds and longer and is also called skin conductance level (SCL). The second is called phasic activity, which changes more rapidly in response to stimuli, on the scale of less than one second, and thus provides a measure of an individual's response to specific affective stimuli (also called skin conductance response, or SCR). It is important to separate these two components of EDA prior to analysis (Boucsein et al. 2012). I used Biopac's *AcqKnowledge* software to separate the components and analyze them separately.

We followed the recommendation to compute EDA levels by taking the integral, or the area under the curve for both the tonic and phasic components (Boucsein et al. 2012; Lajante et al. 2012). The "curve" refers to the time series of each individual participants' EDA level. This measure therefore represents the overall level of physiological arousal by incorporating both the height and the duration of changes in EDA. The unit of electrical activity is a microsiemen (μS); the integral results in units of $\mu S \times$ seconds. I set a threshold of .05 μS to detect SCRs in the phasic data, again following standard recommendations (Lajante et al. 2012).

After recording a 60-second baseline of electrodermal activity, participants began the study by watching one of the two videos used in Study 1 to manipulate mind perception in robots. I bolstered this mind perception manipulation by providing additional written information explaining exactly what was meant by a mind, as in Study 3. Participants were told that minds consist of autonomy and emotional experience. Those in the “mind” condition were further told that robots have both the capacity to think and act with autonomy, without being pre-programmed by humans, and to understand and express emotions, while those in the “no mind” condition were told that robots have neither capacity. Electrodermal activity was recorded throughout this manipulation as a measure of participants’ physiological arousal.

Participants then reported how useful and competent they thought this kind of robot would be, how much they thought such robots could understand what they are thinking and feeling, and how much sympathy they thought robots would have for them if they were suffering.

Results and Discussion

Physiological reactions. The tonic component of EDA, or overall skin conductance level, was lower in the “mind” condition than in the “no mind” condition (M = 353.88 vs. M = 280.25, $t(81) = 1.85$, $p = .067$). The phasic component, or specific skin conductance responses, was not significantly different across conditions (M = -.07 vs. M = .23, $t(81) = 1.38$, $p = .173$). This result is consistent with the fact that the manipulation did not involve specific events or moments that were strongly affective and would thus be expected to elicit specific skin conductance responses; instead, since the manipulation involved watching a short video lecture and reading a description of robot abilities, the slow-changing tonic component of EDA is a more appropriate measure of overall physiological arousal during the manipulation.

Usefulness. Perceived competence and usefulness had high internal reliability ($\alpha = .85$) and were averaged to create a measure of perceived usefulness, which was higher when in the mind condition (M = 6.83) than the no mind condition (M = 4.00, $t(81) = 3.29$, $p = .013$).

Empathy. Having a mind made robots seem more sympathetic to human suffering (M = 3.89 vs. M = 2.56, $t(81) = 2.88$, $p = .005$), but had no effect on robots' perceived ability to take humans' perspective (M = 4.55 vs. M = 4.29, $t(81) = .54$, $p = .589$).

These results largely replicate prior studies using an in-person lab setting and student population. The failure to replicate the finding that mind perception increases robots' ability to take humans' perspective is unexpected and may reflect the fact that the sample size of this study is smaller than in the previous study. Data collection for this study is ongoing in order to increase statistical power. The physiological reactions measured in this study nicely complement the self-reported findings in previous studies to further support the notion that consumers are more comfortable with human-like robots seen as having human-like minds.

GENERAL DISCUSSION

Social robots have the potential to revolutionize multiple sectors of the economy. They are already being deployed in many ways, from working in retail and service jobs, to assisting the elderly and disabled, to augmenting the productivity of human workers in healthcare, education, and more. The potential for these robots to create value for consumers and firms largely depends on them being seen as useful and on humans feeling comfortable around them, and I know very little about what impacts these variables. In study 1, I first replicated the finding that robots with highly but imperfectly human-like physical appearances elicit more negative reactions than robots with low or moderate physical human-likeness and showed that perfect human-likeness improves reactions but that there is still much more room for improvement relative to actual humans. In studies 2 – 5, I then showed that increasing the perception that highly human-like robots can have human-like minds can improve consumer reactions to such robots in terms of increased comfort (self-reported and physiological), perceived usefulness, willingness to visit stores and restaurants where robots are employed, and interest in donating to organizations that promote their development. Furthermore, while mind perception increased perceived usefulness of both moderately and highly human-like robots, it actually decreased comfort with moderately human-like robots, supporting the role of schema congruity in shaping affective reactions to robots.

Limitations and Future Directions

Given that consumers will be actually interacting with robots in consumption contexts, the most important limitation of this research is that participants were not interacting with robots in person, but instead saw pictures or videos of robots interacting with other people. An ongoing field study is currently being conducted to address this limitation, in collaboration with a firm whose goal is to produce perfectly human-like robots called Sanctuary AI. In this study, members of the public will be able to interact with a highly human-like robot in an office environment. I manipulate the robots' autonomy by either allowing it to proactively start and lead conversations (high autonomy) or only respond to questions posed to it (low autonomy). I manipulate the robots' emotional abilities by either allowing it to detect and respond to the emotions of the person interacting with it using affective computing technologies (high emotion) or not allowing it to do so (low emotion). I will measure comfort both via self-report surveys after the interaction and by automatically coding participants' facial expressions during the interaction. The robot will also ask for donations to a pro-robot charity such as the American Society for the Prevention of Cruelty to Robots; this will provide a measure of the robot's usefulness at persuading humans. This study will be complete by late May 2019.

In the present research I did not examine different potential applications of robots in greater detail, instead focusing mainly on consumers' general reactions. In Study 2, I did find that the effects mind perception had the same effects on consumers' comfort with robots as a store clerk and as a waitress, suggesting that our effects are likely to be stable across different contexts and applications. Furthermore, in study 3 I found that the effects of mind perception were similar (although slightly stronger) in a task that involved direct social interaction with the robot (i.e., nurse vs. hospital administrator). Nevertheless, it would be worthwhile to further investigate whether there are specific applications in which human-likeness and mind perception matter more or less to consumers. For example, while hospital administrators do of course interact less directly with patients less than nurses do, there are other jobs in which social interaction plays an even smaller role, and for those tasks I may see that mind perception plays a smaller role.

In terms of further theory building, future research should explore the lay psychology of mind perception in more detail. Specifically, our results show that consumers' reactions to social robots improve when they think that robots have a human-like mind, which may suggest that consumers are, philosophically speaking, "lay dualists" whose default belief is that the mind in something immaterial and not feasible for robots to have. Alternatively, consumers may initially be open to a materialist view of the mind and simply be unwilling to grant robots a complete mind for other reasons. One such reason might be what we call "species-ism," meaning that humans may discriminate against robots even if the robots appear to have perfectly human-like appearances and minds, simply because they are not members of the biological human species. Research opportunities may therefore exist in terms of extending research on other forms of discrimination such as sexism and racism into the domain of human-robot interaction. In general, studying the relationships between lay theories of mind and reactions to robots in greater depth seems to be a fruitful path forward.

Conclusion

Social robots are the fastest growing segment of the robotics market, but research on how consumers react to them is scarce. As these robots begin to truly blur the line between human and machine, this Essay provides guidance for the creators, employers, and marketers of human-like robots to navigate out of the uncanny valley. Both physical and mental human-likeness need to be considered as factors that shape social robots' value for consumers and firms. Incorporating both of these factors is also important for future research in anthropomorphism, mind perception, and human-robot interaction.

CHAPTER 5: CONCLUSION

Several broad conclusions can be drawn from the findings presented in this Dissertation. The first is that a technology's human-likeness is a crucial determinant of the technology's adoption and of consumers' perceptions of the technology. Consumer research on product anthropomorphism has explored superficial aspects of human-likeness such as whether a car appears to be smiling or frowning (Aggarwal and McGill 2007), but the advent of sophisticated AI requires a much more nuanced conceptualization of what it means for a technology to be human-like. Essay 1 provided this conceptualization, and Essays 2 and 3 demonstrated that the three dimensions of human-likeness (physical, cognitive, and emotional) are robust predictors of consumer perceptions and adoption of AI. While it can thus be concluded that human-likeness is an important variable to understand and control, many opportunities remain for future research to moderators of human-likeness effects, how the three dimensions interact with each other in different contexts, and whether there are non-linear effects of human-likeness.

A second conclusion is that the creators and marketers of AI products will likely face a frequent challenge in balancing their products' perceived usefulness and consumers' comfort with the products. Essays 2 and 3 both demonstrated this tension: increasing the human-likeness of a technology can make it seem more useful while also making consumers less comfortable with it. Comfort (or affective reactions more broadly) with new technologies has not been a feature of the most prominent existing models of technology adoption (Venkatesh et al. 2003), and the extent to which comfort is a relevant consideration in the adoption of most technologies is far different in the context of human-like AI. As machines truly acquire physical, cognitive, and emotional human-likeness, fundamental notions about what it means to be human and what separates us from machines are challenged, thus creating the potential for a qualitatively different form of discomfort than is relevant to non-human-like technologies. Ensuring that AI has a positive impact on consumers, firms, and society therefore demands a careful consideration not only of the technology's usefulness, but also of how the technology's existence and use might affect consumers on a more affective level.

A third conclusion is that maximizing this potential value while minimizing the risks requires not only a technical approach to the technology itself, but also a psychological approach to the consumers of the technology. Both B2B and B2C applications of AI will be of extremely limited value if managers and consumers respectively are not comfortable with the technology and/or do not perceive it as useful. As demonstrated in this Dissertation, perceived usefulness is not exclusively a function of the technology's objective performance, as consumers continue to prefer relying on humans rather than algorithms even when they know the algorithm performs better. Perceived usefulness is indeed distinct from the "objective usefulness" as defined by the technology's capabilities, and the distance between perceived and objective usefulness is determined by psychological variables. Comfort is more obviously psychological and independent from AI's technical features and benefits. Understanding the determinants of both perceived usefulness comfort via rigorous experimentation will improve both firms' own confidence in their products' value and their ability to sell the products by using the resulting empirical evidence as a marketing asset. This Dissertation provides an illustration of how this evidence can be created and used to provide value to the producers and consumers of AI technologies.

This Dissertation also offers several practical contributions, in the form of specific marketing implications. Table 8 summarizes the practical marketing recommendations that can be made on the basis of specific empirical findings in this Dissertation. These recommendations illustrate the value that firms can obtain by taking a theory-driven experimental approach to developing and advertising their AI-enabled products and services.

Marketing Task	Marketing Recommendations	Based on Finding
New Product Development	Develop and employ robots with low or moderate (rather than high) human-likeness.	Human-likeness decreases consumers' comfort with robots (Essay 3, Study 1)
	When employing robots with high human-likeness, ensure they have emotion recognition capabilities and sufficient autonomy to pro-actively lead conversations with humans.	Endowing highly human-like robots with emotional abilities and autonomy improves consumers' comfort with and perceived usefulness of such robots (Essay 3, Studies 2–5).
Advertising	When employing robots with high human-likeness, advertise such robots using physicalist (vs. dualist) descriptions of their mental abilities.	Believing that minds are physical things which robots can have increases consumers' comfort with and perceived usefulness of highly human-like robots (Essay 3, Studies 2, 4, and 5).
	When advertising algorithm-based products or services, emphasize the quantitative aspects of the task for which the algorithm is being used.	Emphasizing a tasks' quantitative elements makes it seem more objective and increases consumers' trust in and use of algorithms for the task (Essay 2, Studies 3, 5, and 6).
	When advertising algorithm-based products or services for tasks normally seen as subjective, emphasize algorithms' affective abilities.	Increasing algorithms' affective human-likeness increases comfort with and perceived usefulness of algorithms for subjective tasks (Essay 2, Study 6).

Table 8. Marketing recommendations based on Dissertation findings.

This Dissertation also raises some deeper questions that warrant further study. First, how can the notion of human-likeness developed here inform our understanding of human consumers? The dominant model of consumers in the academic consumer behavior literature for many years was a purely cognitive information-processing model, often likened to a computer (Bettman 1970; Howard and Sheth 1969). Many scholars have recognized the need to update those models to include more affective components, such as the role of emotions in consumer behavior (Holbrook and Hirschman 1982; Pham 1998). More recently, there has also been a growing awareness of the importance of physicality in shaping consumer behavior, reflected in the embodied cognition literature (Adam and Galinsky 2012; Krishna 2012). Human-likeness as defined in this Dissertation contains three components that correspond to these three paradigms of the broader consumer behavior literature: cognitive human-likeness, corresponding to the first paradigm of information processing models of consumer behavior; affective human-likeness, corresponding to the second paradigm of feelings and emotions in consumer behavior; and physical human-likeness, corresponding to the third paradigm of embodied cognition.

This convergence suggests that research on AI and research on consumer psychology can complement and learn from each other. Attempts to create human-like machines has required computer scientists and engineers to define what it means to be human-like; similarly, attempts to understand consumer behavior have led behavioral scientists to build up their own understanding of how humans learn, think, feel, and decide. However, while consumer behavior researchers have largely accepted the importance of feelings and physicality in addition to cognition, the three paradigms of consumer behavior research remain largely siloed from one another (Peracchio, Luce, and McGill 2014). This separation highlights the need for a more integrative approach, one that recognizes and models the simultaneous and interactive influences of cognitive, affective, and physical processes in shaping consumer behavior.

A second question is whether there are historical parallels to the rise of AI that can shed light on how it is likely to be adopted and used in society and what the limitations of those parallels might be. Several existing technologies have been profoundly transformational, including the printing press, the electric light, the transistor, and the Internet. It seems likely that both the perceived usefulness of these technologies as well as consumers' comfort using them had major effects on their rate and pattern of adoption, as I expect will be the case with AI. A deeper historical analysis can undoubtedly reveal additional dynamics and forces that shaped the adoption of these transformational technologies which may also be relevant today. One apparent factor distinguishing AI from any prior technology, however, is human-likeness, which AI has but which no other technology does. As demonstrated throughout this Dissertation, human-likeness has strong effects on consumers' perceptions and adoption of AI technologies. These effects are largely mediated by usefulness and comfort, suggesting intuitive parallels between AI and existing technologies in terms of the proximate antecedents of adoption, but important differences in terms of the more ultimate causes of adoption.

Finally, a pressing open question facing those interested in AI adoption is how this technology should be governed in order to maximize its value and minimize its risks. As noted throughout this Dissertation, AI has the potential to create unprecedented value for consumers and firms, but also to profoundly disrupt the economic, political, and psychological status quo. Virtually none of organizations at the forefront of developing AI are doing so with the goal of benefiting all of humanity (with the possible exception of OpenAI) – instead, they are either national governments hoping that AI will make their countries more competitive and powerful, or technology companies using AI to increase their profitability. Neither of these motivations are likely to be beneficial for the majority of humanity. Instead, the benefits will accrue to the citizens of the successful countries and the shareholders of the successful companies. In the case of certain governments (i.e., in China) the benefits may not even accrue to all citizens, but only to "preferred" ethnic groups (Mozur 2019) These motivations also suggest specific risks including exacerbated global income inequality, an oligopolistic global market structure, increased totalitarianism (i.e., by tracking and controlling citizen behavior online and offline), and an arms race to develop AI that could develop into full-fledged war (Dafoe 2018). These issues clearly move far beyond the realm of marketing, but their importance is clear, and they demand an interdisciplinary, multi-national, public-private effort to develop norms, institutions, and rules to govern AI's development. One role that marketing research can play in this effort is to study how to increase the general public's prioritization of AI governance as an important societal goal, and specifically how this can be done in different cultural contexts, in order to increase the likelihood of such governance ultimately developing. Maximizing the potential value of this stream of research may therefore require a shift from understanding adoption of AI itself to understanding – and shaping – public attitudes towards the development, use, and governance of AI by powerful nations and corporations.

REFERENCES

Acemoglu, Daron and Pascual Restrepo (2018), "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 108(6), 1488–1542.

Adam, Hajo and Adam D. Galinsky (2012), "Enclothed Cognition," *Journal of Experimental Social Psychology*, 48(4), 918–25.

Aggarwal, Pankaj and Ann L. McGill (2007), "Is That Car Smiling at Me? Schema Congruity as a Basis for Evaluating Anthropomorphized Products," *Journal of Consumer Research*, 34(4), 468–79.

André, Quentin, Ziv Carmon, Klaus Wertenbroch, Alia Crum, Douglas Frank, William Goldstein, Joel Huber, Leaf van Boven, Bernd Weber, and Haiyang Yang (2018), "Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data," *Customer Needs and Solutions*, 5(1–2), 28–37.

Armstrong, J. Scott (1980), "The Seer-Sucker Theory: The Value of Experts in Forecasting," *Technology Review*, 82(7), 16–24.

Avnet, Tamar, Michel Tuan Pham, and Andrew Stephen (2012), "Consumers' Trust in Feelings as Information," *Journal of Consumer Research*, 39(4), 720–35.

Balconi, Michela, Luciana Falbo, and Veronica Anna Conte (2012), "BIS and BAS Correlates with Psychophysiological and Cortical Response Systems during Aversive and Appetitive Emotional Stimuli Processing," *Motivation and Emotion*, 36(2), 218–31.

Barrett, Lisa Feldman, Batja Mesquita, Kevin N. Ochsner, and James J. Gross (2006), "The Experience of Emotion," *Annual Review of Psychology*, 58(1), 373–403.

Bartneck, Christoph, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita (2009), "My Robotic Doppelgänger - A Critical Look at the Uncanny Valley," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 269–76.

Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15(5), 215–27.

Baumeister, Roy F., Kathleen D. Vohs, Jennifer L. Aaker, and Emily N. Garbinsky (2013), "Some Key Differences between a Happy Life and a Meaningful Life," *Journal of Positive Psychology*.

Bear, Adam and David G. Rand (2016), "Intuition, Deliberation, and the Evolution of Cooperation," *Proceedings of the National Academy of Sciences*, 113(4), 936–41.

Beer, Jenay M, Arthur D Fisk, and Wendy A Rogers (2014), "Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction," *Journal of Human-Robot Interaction*, 3(2), 74–102.

Berg, Andrew, Edward Buffie, and Luis-Felipe Zanna (2016), "Robots, Growth, and Inequality," *Finance and Development*, 32(September), 10–13.

Bettman, James (1973), "Perceived Risk and Its Components: A Model and Empirical Test," *Journal of Marketing*, 10(2), 184–90.

Bettman, James R. (1970), "Information Processing Models of Consumer Behavior," *Journal of Marketing Research*, 7(3), 370–76.

Bishop, Christopher M. (2006), *Pattern Recognition and Machine Learning*, New York: Springer.

Blattberg, Robert C and Stephen J Hoch (1990), "Database Models and Managerial Intuition: 50% Model + 50% Manager," *Management Science*, 36(8), 887–99.

Bostrom, Nick (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press.

Boucsein, Wolfram, Don C. Fowles, Sverre Grimnes, Gershon Ben-Shakhar, Walton T. Roth, Michael E. Dawson, and Diane L. Filion (2012), "Publication Recommendations for Electrodermal Measurements," *Psychophysiology*, 49(8), 1017–34.

Bradshaw, Tim (2015), "Scientists and Investors Warn on AI," *Financial Times*.

Bucklin, Randolph E., Donald R. Lehmann, and John D.C. Little (1998), "From Decision Support to Decision Automation: A 2020 Vision," *Marketing Letters*, 9(3), 235–46.

Bughin, Jacques, Jeongmin Seong, James Manyika, and Raoul Joshi (2018), *Notes From The AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute.

Buolamwini, Joy and Timnit Gebru (2018), "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91.

Business Insider (2015), "Growth Statistics For Robots Market," *Business Insider*, http://www.businessinsider.com/growth-statistics-for-robots-market-2015-2.

Business Wire (2017), "$5.43 Billion Consumer Robot Market 2017 - Industry Trends, Opportunities and Forecasts to 2023," *Business Wire*, https://www.businesswire.com/news/home/20171219005736/en/5.43-Billion-Consumer-Robot-Market-2017.

Byrnes, James P., David C. Miller, and William D. Schafer (1999), "Gender Differences in Risk Taking: A Meta-Analysis," *Psychological Bulletin*, 125(3), 367–83.

Cellan-Jones, Rory (2014), "Stephen Hawking Warns Artificial Intelligence Could End Mankind," *BBC News*.

Chopra, K. and W. A. Wallace (2003), "Trust in Electronic Environments," in *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*, 1–10.

Clerwall, Christer (2014), "Enter the Robot Journalist," *Journalism Practice*, 8(5), 519–31.

Cummings, M.L. (2017), *Artificial Intelligence and the Future of Warfare*, Chatham House.

Dafoe, Allan (2018), *AI Governance: A Research Agenda*.

Dass, Camillia (2017), "Asia's First Retail Robot Launched to Help Relieve Manpower Shortages," *The Straits Times*.

Davis, Fred, Richard Bagozzi, and Peter Warshaw (1989), "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science*, 35(8), 982–1003.

Dawes, Robyn M, David Faust, and Paul E Meehl (1989), "Clinical versus Actuarial Judgment.," *Science*, 243(4899), 1668–74.

Deahl, Dani (2018), "How AI-Generated Music Is Changing the Way Hits Are Made," *The Verge*.

Deci, Edward L. and Richard M. Ryan (2000), "The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry*, 11(4), 227–68.

Dennett, Daniel C (1997), "Consciousness in Human and Robot Minds," in *Cognition, Computation, and Consciousness*, ed. M. Ito, Y. Miyashita, and E. Rolls, Oxford: Oxford University Press, 17–29.

Dhar, Ravi and Klaus Wertenbroch (2000), "Consumer Choice Between Hedonic and Utilitarian Goods," *Journal of Marketing Research*, 37(1), 60–71.

Diab, Dalia L., Shuang-Yueh Pui, Maya Yankelevich, and Scott Highhouse (2011), "Lay Perceptions of Selection Decision Aids in US and Non-US Samples," *International Journal of Selection and Assessment*, 19(2), 209–16.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2016), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science*, 64(3), 1155–70.

Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey (2014), "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err," *Journal of Experimental Psychology: General*, 143(6), 1–13.

Dijksterhuis, Ap, Maarten W. Bos, Loran F. Nordgren, and Rick B. Van Baaren (2006), "On Making the Right Choice: The Deliberation-without-Attention Effect," *Science*, 311(5763), 1005–7.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein, and Kaiya Liu (2011), "Partitioning Default Effects: Why People Choose Not to Choose," *Journal of Experimental Psychology: Applied*, 17(4), 332–41.

Domingos, Pedro (2012), "A Few Useful Things to Know about Machine Learning," *Communications of the ACM*, 55(10), 78.

Dosilovic, Filip Karlo, Mario Brcic, and Nikica Hlupic (2018), "Explainable Artificial Intelligence: A Survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 0210–15.

Dujmovic, Jurica (2017), "What's Holding Back Artificial Intelligence? Americans Don't Trust It," *MarketWatch*, https://www.marketwatch.com/story/whats-holding-back-artificial-intelligence-americans-dont-trust-it-2017-03-30.

Eisenhardt, Kathleen M. (1989), "Agency Theory : An Assessment and Review," *Academy of*

Management Review, 14(1), 57–74.

Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun (2017), "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature*, 542(7639), 115–18.

Faraji-Rad, Ali, Shiri Melumad, and Gita Venkataramani Johar (2017), "Consumer Desire for Control as a Barrier to New Product Adoption," *Journal of Consumer Psychology*, 27(3), 347–54.

Ferrari, Francesco, Maria Paola Paladino, and Jolanda Jetten (2016), "Blurring Human–Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness," *International Journal of Social Robotics*, 8(2), 287–302.

Fishbein, Martin and Icek Ajzen (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*, Reading, MA: Addison-Wesley.

Fiske, Susan T., Amy J. C. Cuddy, and Peter Glick (2007), "Universal Dimensions of Social Cognition," *Trends in Cognitive Sciences*, 11(2), 77–83.

Folkes, Valerie S. (1988), "The Availability Heuristic and Perceived Risk," *Journal of Consumer Research*, 15(1), 13–23.

Frith, Chris (2009), "Role of Facial Expressions in Social Interactions," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3453–58.

Furman, Jason and Robert Seamans (2018), *AI and the Economy*, NBER Working Paper Series, Cambridge, MA.

Galinsky, Adam D., Gillian Ku, and Cynthia S. Wang (2005), "Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination," *Group Processes and Intergroup Relations*, 8(2 SPEC. ISS.), 109–24.

Gibbs, Samuel (2016a), "Five In-Home Robots That Could Change Your Life," *The Guardian*.

——— (2016b), "Google AI Project Writes Poetry Which Could Make a Vogon Proud," *The Guardian*.

Gigerenzer, Gerd and Henry Brighton (2009), "Homo Heuristicus: Why Biased Minds Make Better Inferences," *Topics in Cognitive Science*, 1(1), 107–43.

Goasduff, Laurence (2017), "Emotion AI Will Personalize Interactions," *Gartnet*, https://www.gartner.com/smarterwithgartner/emotion-ai-will-personalize-interactions/.

Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton (2013), "Speech Recognition with Deep Recurrent Neural Networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 6645–49.

Gray, Heather M, Kurt Gray, and Daniel M Wegner (2007), "Dimensions of Mind Perception," *Science*, 619(February), 10–11.

Gray, Kurt and Daniel M. Wegner (2012), "Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley," *Cognition*, 125(1), 125–30.

Greene, Joshua and Jonathan Cohen (2012), "For The Law, Neuroscience Changes Nothing and Everything," in *Oxford Handbook of Neuroethics*, ed. Judy Illes and Barbara Sahakian, Oxford: Oxford University Press.

Greifeneder, Rainer, Herbert Bless, and Michel Tuan Pham (2011), "When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review," *Personality and Social Psychology Review*, 15(2), 107–41.

Grove, William M., David H. Zald, Boyd S. Lebow, Beth E. Snitz, and Chad Nelson (2000), "Clinical versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 12(1), 19–30.

Grove, William M, David H Zald, Boyd S Lebow, Beth E Snitz, and Chad Nelson (2000), "Clinical versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 12(1), 19–30.

Gustafson, Pierre (1998), "Gender Differences in Risk Perception: Theoretical and Methodological Perspectives," *Risk Analysis*, 18(6), 805–11.

Hanson, David (2005), "Expanding the Aesthetic Possibilities for Humanoid Robots," in *IEEE-RAS International Conference on Humanoid Robots*, 24–31.

Harnad, Stevan (1991), "Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem," *Minds and Machines*, 1(1), 43–54.

Haslam, Nick (2006), "Dehumanization: An Integrative Review," *Personality and Social Psychology Review*, 10(3), 252–64.

Haslam, Nick, Yoshihisa Kashima, Stephen Loughnan, Junqi Shi, and Caterina Suitner (2008), "Subhuman, Inhuman, and Superhuman: Contrasting Humans with Nonhumans in Three Cultures," *Social Cognition*, 26(2), 248–58.

Haslam, Nick and Steve Loughnan (2014), "Dehumanization and Infrahumanization," *Annual Review of Psychology*, 65, 399–423.

Herremans, Dorien, David Martens, and Kenneth Sörensen (2014), "Dance Hit Song Prediction," *Journal of New Music Research*, 43(3), 291–302.

Higgins, E. Tory and William S. Rholes (1978), "'Saying Is Believing': Effects of Message Modification on Memory and Liking for the Person Described," *Journal of Experimental Social Psychology*, 14(4), 363–78.

Ho, Cc, Kf MacDorman, and Zad Pramono (2008), "Human Emotion and the Uncanny Valley: A GLM, MDS, and Isomap Analysis of Robot Video Ratings," *Human Robot Interactions*, 8(January), 169–76.

Holbrook, Morris B. and Elizabeth C. Hirschman (1982), "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun," *Journal of Consumer Research*, 9(2), 132–40.

Howard, John. A. and Jagdish. N. Sheth (1969), "A Theory of Buyer Behavior," *John Wiley& Sons*.

Hoynes, Hilary and Jesse Rothstein (2019), *Universal Basic Income in the US and Advanced Countries*, NBER Working Paper Series.

Hughes, James J (2014), "A Strategic Opening for a Basic Income Guarantee in the Global Crisis Being Created by AI, Robots, Desktop Manufacturing and BioMedicine," *Journal of*

Evolution and Technology, 24(1), 45–61.

Inbar, Yoel, Jeremy Cone, and Thomas Gilovich (2010), "People's Intuitions About Intuitive Insight and Intuitive Choice," *Journal of Personality and Social Psychology*, 99(2), 232–47.

Inglehart, Ronald and Pippa Norris (2017), "Trump and the Populist Authoritarian Parties: The Silent Revolution in Reverse," *Perspectives on Politics*, 15(2), 443–54.

Jacoby, Jacob and Leon B. Kaplan (1972), "The Components of Perceived Risk," *ACR Special Volumes*, (January), 382–93.

Johnson, Devon and Kent Grayson (2005), "Cognitive and Affective Trust in Service Relationships," *Journal of Business Research*, 58(4), 500–507.

Joshi, A K (1991), "Natural Language Processing," *Science*, 253(5025), 1242–49.

Kahn, Jeremy (2018), "Just How Shallow Is the Artificial Intelligence Talent Pool?," *Bloomberg*, https://www.bloomberg.com/news/articles/2018-02-07/just-how-shallow-is-the-artificial-intelligence-talent-pool.

Kätsyri, Jari, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala (2015), "A Review of Empirical Evidence on Different Uncanny Valley Hypotheses: Support for Perceptual Mismatch as One Road to the Valley of Eeriness," *Frontiers in Psychology*, 6(March), 32–50.

Kelleher, Kevin (2018), "37% of Tech Experts Worry AI Will Make Life Worse by 2030," *Fortune*, http://fortune.com/2018/12/14/artificial-intelligence-impact-ai-research-survey-experts/.

Keltner, Dacher and Jonathan Haidt (1999), "Social Functions of Emotions at Four Levels of Analysis," *Cognition & Emotion*, 13(5), 505–21.

Khatchadourian, Raffi (2015), "We Know How You Feel," *The New Yorker*, 50–59.

Kodra, Evan, Thibaud Senechal, Daniel McDuff, and Rana El Kaliouby (2013), "From Dials to Facial Coding: Automated Detection of Spontaneous Facial Expressions for Media Research," in *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 15–21.

Komiak, Sherrie and Izak Benbasat (2006), "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *Management Information Systems Quarterly*, 30(4), 941–60.

Kosinski, M., D. Stillwell, and T. Graepel (2013), "Private Traits and Attributes Are Predictable from Digital Records of Human Behavior," *Proceedings of the National Academy of Sciences*, 110(15), 5802–5.

Krasnianski, Adriana (2015), "Meet Ross, the IBM Watson-Powered Lawyer," *PSFK*, http://www.psfk.com/2015/01/ross-ibm-watson-powered-lawyer-legal-research.html.

Krishna, Aradhna (2012), "An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior," *Journal of Consumer Psychology*, 22(3), 332–51.

Kunda, Ziva (1990), "The Case for Motivated Reasoning.," *Psychological Bulletin*, 108(3), 480–98.

Lafrance, Adrienne (2015), “Self-Driving Cars Could Save 300,000 Lives Per Decade in America,” *The Atlantic*, September.

Lajante, Mathieu, Olivier Droulers, Thibaut Dondaine, and David Amarantini (2012), “Opening the ‘Black Box’ of Electrodermal Activity in Consumer Neuroscience Research,” *Journal of Neuroscience, Psychology, and Economics*, 5(4), 238–49.

Lang, Peter, Mark Greenwald, Margaret Bradley, and Alfonso Hamm (1993), “Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions,” *Psychophysiology*, 30(3), 261–73.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015), “Deep Learning,” *Nature*, 521(7553), 436–44.

Lee, J. D. and K. A. See (2004), “Trust in Automation: Designing for Appropriate Reliance,” *Human Factors*, 46(1), 50–80.

Leotti, Lauren A., Sheena S. Iyengar, and Kevin N. Ochsner (2010), “Born to Choose: The Origins and Value of the Need for Control,” *Trends in Cognitive Sciences*, 457–63.

Leung, Eugina, Gabrielle Paolacci, and Stefano Puntoni (2018), “Man versus Machine: Resisting Automation in Identity-Based Consumer Behavior,” *Journal of Marketing Research*, 55(6), 818–31.

Li, Xin, Traci J. Hess, and Joseph S. Valacich (2008), “Why Do We Trust New Technology? A Study of Initial Trust Formation with Organizational Information Systems,” *The Journal of Strategic Information Systems*, 17(1), 39–71.

Liu, Bing (2010), “Sentiment Analysis and Subjectivity,” in *Handbook of Natural Language Processing*, ed. Nitin Indurkhya and Fred Damerau, Boca Raton, FL: CRC Press, 1–38.

Loewenstein, George F., Christopher K. Hsee, Elke U. Weber, and Ned Welch (2001), “Risk as Feelings,” *Psychological Bulletin*, 127(2), 267–86.

Logg, Jennifer M (2017), “Theory of Machine: When Do People Rely on Algorithms?”

Logg, Jennifer, Julia Minson, and Don Moore (2019), “Algorithm Appreciation: People Prefer Algorithmic To Human Judgment,” *Organizational Behavior and Human Decision Processes*, 151.

Lohr, Steve (2018), “Are Robots Coming for Your Job? Eventually, Yes,” *The New York Times*.

Loken, Barbara (2006), “Consumer Psychology: Categorization, Inferences, Affect, and Persuasion,” *Annual Review of Psychology*, 57(1), 453–85.

Loughnan, Stephen and Nick Haslam (2007), “Animals and Androids: Implicit Associations between Social Categories and Nonhumans,” *Psychological Science*, 18(2), 116–22.

MacDorman, Karl F., Robert D. Green, Chin Chang Ho, and Clinton T. Koch (2009), “Too Real for Comfort? Uncanny Responses to Computer Generated Faces,” *Computers in Human Behavior*, 25(3), 695–710.

Malmendier, Ulrike and Geoffrey Tate (2005), “CEO Overconfidence and Corporate Investment,” *Journal of Finance*, 60(6), 2661–2700.

Manning, Susan Karp and Maria Parra Melchiori (1974), “Words That Upset Urban College

Students: Measured with Gsrs and Rating Scales," *Journal of Social Psychology*, 94(2), 305–6.

Manyika, James and Jacques Bughin (2018), *The Promise and Challenge of the Age of Artificial Intelligence*, McKinsey Global Institute.

Manyika, James, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst (2017), *A Future That Works: Automation, Employment, and Productivity*, Mckinsey Global Institute.

Manyika, James, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko, and Saurabh Sanghvi (2017), *Jobs Lost, Jobs Gained: What the Future of Work Will Mean for Jobs, Skills, and Wages*, McKinsey Global Institute.

Mathur, Maya B. and David B. Reichling (2016), "Navigating a Social World with Robot Partners: A Quantitative Cartography of the Uncanny Valley," *Cognition*, 146(8), 22–32.

Matz, Sandra, Michael Kosinski, Gideon Nave, and David Stillwell (2017), "Psychological Targeting as an Effective Approach to Digital Mass Persuasion," *Proceedings of the National Academy of Sciences*, 114(48), 12714–19.

McDuff, Daniel, Rana El Kaliouby, Thibaud Senechal, May Amr, Jeffrey F. Cohn, and Rosalind Picard (2013), "Affectiva-Mit Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected 'in-the-Wild,'" in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 881–88.

McDuff, Daniel, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby (2016), "AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit," in *CHI Conference on Human Factors in Computing Systems*, 3723–26.

Meehl, Paul E. (1954), *Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis: University of Minnesota Press.

Metz, Cade (2017), "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," *The New York Times*.

Meyer, Jenny (2011), "Workforce Age and Technology Adoption in Small and Medium-Sized Service Firms," *Small Business Economics*, 37(3), 305–24.

Meyers-Levy, Joan and Alice M. Tybout (2002), "Schema Congruity as a Basis for Product Evaluation," *Journal of Consumer Research*, 16(1), 39–54.

Mitchell, Wade J., Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman (2011), "A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley," *i-Perception*, 2(1), 10–12.

Moore, Don A. and Paul J. Healy (2008), "The Trouble with Overconfidence," *Psychological Review*, 115(2), 502–17.

Moore, Gary C. and Izak Benbasat (1991), "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research*, 2(3), 192–222.

Morales, Andrea C., On Amir, and Leonard Lee (2017), "Keeping It Real in Experimental

Research-Understanding When, Where, and How to Enhance Realism and Measure Consumer Behavior," *Journal of Consumer Research*, 44(2), 465–76.

Mori, Masahiro (1970), "The Uncanny Valley," *Energy*, 7(4), 33–35.

Morse, Nancy C. and Robert S. Weiss (1955), "The Function and Meaning of Work and the Job," *American Sociological Review*, 20(2), 191–209.

Mozur, Paul (2019), "One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority - The New York Times," *The New York Times*.

Müller, Vincent C. and Nick Bostrom (2016), "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in *Fundamental Issues of Artificial Intelligence*, ed. Vincent Muelle, Berlin: Springer International Publishing, 555–72.

Murphy, Kevin P. (2012), *Machine Learning: A Probabilistic Perspective*, Cambridge, MA: MIT Press.

Nguyen, Clinton (2016), "Chinese Restaurants Are Replacing Waiters with Robots," *Business Insider*, http://www.businessinsider.com/chinese-restaurant-robot-waiters-2016-7/.

Nichols, Shaun (2008), "How Can Psychology Contribute to the Free Will Debate?," in *Are We Free? Psychology and Free Will*, ed. John Baer, James Kaufman, and Roy Baumeister, Oxford: Oxford University Press.

Nikolić, Milica, Evin Aktar, Susan Bögels, Cristina Colonnesi, and Wieke de Vente (2018), "Bumping Heart and Sweaty Palms: Physiological Hyperarousal as a Risk Factor for Child Social Anxiety," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 59(2), 119–28.

O'Neil, Cathy (2016), *Weapons of Math Destruction*, New York: Crown.

Oswald, Patricia A. (1996), "The Effects of Cognitive and Affective Perspective Taking on Empathic Concern and Altruistic Helping," *Journal of Social Psychology*, 136(5), 613–23.

Padmavathi, Thimmaiah and Madaiah Krishnamurthy (2017), "Semantic Web Tools and Techniques for Knowledge Organization: An Overview," *Knowledge Organization*, 44(4), 273–90.

Park, Dong Huk, Lisa Anne Hendricks, Zeynep Akata, Anna Rohrbach, Bernt Schiele, Trevor Darrell, and Marcus Rohrbach (2017), *Attentive Explanations: Justifying Decisions and Pointing to the Evidence*, arXiv Preprint.

Parker, Jeffrey R., Donald R. Lehmann, and Yi Xie (2016), "Decision Comfort," *Journal of Consumer Research*, 43(1), 113–33.

Parkinson, Brian (1996), "Emotions Are Social," *British Journal of Psychology*, 87(4), 663–83.

Paulheim, Heiko (2017), "Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods," *Semantic Web*, 8(3), 489–508.

Pavlou, Paul A. and David Gefen (2004), "Building Effective Online Marketplaces with Institution-Based Trust," *Information Systems Research*, 15(1).

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti (2017), "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research," *Journal of*

Experimental Social Psychology, 70, 153–63.

Peracchio, Laura A., Mary Frances Luce, and Ann L. McGill (2014), "Building Bridges for an Interconnected Field of Consumer Research," *Journal of Consumer Research*.

Perera, Rivindu and Parma Nand (2017), "Recent Advances in Natural Language Generation: A Survey and Classification of the Empirical Literature," *Computing and Informatics*, 36(1), 1–32.

Perkins, Anne (2018), "May to Pledge Millions to AI Research Assisting Early Cancer Diagnosis," *The Guardian*, May 20.

Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller (2016), *Artificial Intelligence and Life in 2030*, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel.

Pham, Michel Tuan (1998), "Representativeness, Relevance, and the Use of Feelings in Decision Making," *Journal of Consumer Research*, 25(2), 144–59.

Phillips, Elizabeth, Xuan Zhao, Daniel Ullman, and Bertram F Malle (2018), "What Is Human-like?: Decomposing Robots' Human-like Appearance Using the Anthropomorphic RoBOT (ABOT) Database," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 105–13.

Picard, Rosalind (2011), "Measuring Affect in the Wild," in *Affective Computing and Intelligent Interaction*, ed. Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, Memphis, TN, 6974.

Pratt, Michael G. and B. Ashforth (2003), "Fostering Meaningfulness in Working and at Work," in *Positive Organizational Scholarship: Foundations of a New Discipline*, ed. K.S. Cameron, J.E. Dutton, and R.E. Quinn, San Francisco: Berrett-Koehler Publishers, Inc., 309–27.

Press, Gil (2019), "4 Observations From Recent Surveys About The State-Of-Artificial Intelligence (AI)," *Forbes*, https://www.forbes.com/sites/gilpress/2019/02/28/4-observations-from-recent-surveys-about-the-state-of-artificial-intelligence-ai/#401434114e7a.

Promberger, Marianne and Jonathan Baron (2006), "Do Patients Trust Computers?," *Journal of Behavioral Decision Making*, 19(5), 455–68.

Quackenbush, Casey (2018), "Painting Made by Artificial Intelligence Sells for $432,500," *Time*.

Ritchie, G, R Manurung, H Pain, and A Waller (2007), "A Practical Application of Computational Humour," *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 91–98.

Rodrik, Dani (2018), "Populism and the Economics of Globalization," *Journal of International Business Policy*, 1, 12–33.

Rogers, Everett M. (1976), "New Product Adoption and Diffusion," *Journal of Consumer*

Research, 2(March), 290–301.

Rosenthal-Von Der Pütten, Astrid M. and Nicole C. Krämer (2014), "How Design Characteristics of Robots Determine Evaluation and Uncanny Valley Related Responses," *Computers in Human Behavior*, 36, 422–39.

Rosenthal-von der Pütten, Astrid M., Frank P. Schulte, Sabrina C. Eimler, Sabrina Sobieraj, Laura Hoffmann, Stefan Maderwald, Matthias Brand, and Nicole C. Krämer (2014), "Investigations on Empathy towards Humans and Robots Using FMRI," *Computers in Human Behavior*, 33, 201–12.

Rosso, Brent D., Kathryn H. Dekas, and Amy Wrzesniewski (2010), "On the Meaning of Work: A Theoretical Integration and Review," *Research in Organizational Behavior*, 30, 91–127.

Russell, Stuart J. and Peter Norvig (2009), *Artificial Intelligence: A Modern Approach, 3rd Edition*, London: Pearson.

Saygin, Ayse Pinar, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith (2012), "The Thing That Should Not Be: Predictive Coding and the Uncanny Valley in Perceiving Human and Humanoid Robot Actions," *Social Cognitive and Affective Neuroscience*, 7(4), 413–22.

Scheinkman, José A and Wei Xiong (2003), "Overconfidence and Speculative Bubbles," *Journal of Political Economy*, 111(6), 1183–1220.

Searle, John Rogers (1995), *The Rediscovery of the Mind*, Cambridge, MA: MIT Press.

Selwyn, Neil, Stephen Gorard, John Furlong, and Louise Madden (2003), "Older Adults' Use of Information and Communications Technology in Everyday Life," *Ageing and Society*, 23(5), 561–82.

Shariff, Azim, Jean-François Bonnefon, and Iyad Rahwan (2017), "Psychological Roadblocks to the Adoption of Self-Driving Vehicles," *Nature Human Behaviour*, 1(10), 694.

Shiv, Baba and Alexander Fedorikhin (1999), "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making," *Journal of Consumer Research*, 26(3), 278–92.

Simon, Matt (2015), "This Incredible Hospital Robot Is Saving Lives. Also, I Hate It," *WIRED*, https://www.wired.com/2015/02/incredible-hospital-robot-saving-lives-also-hate/.

Simonite, Tom (2014), "IBM Watson's Plan to End Human Doctors' Monopoly on Medical Know-How," *MIT Technology Review*.

Simonsohn, Uri (2018), "Two Lines: A Valid Alternative to the Invalid Testing of U-Shaped Relationships With Quadratic Regressions," *Advances in Methods and Practices in Psychological Science*, 1(4), 538–55.

Singer, Tania and Olga M. Klimecki (2014), "Empathy and Compassion," *Current Biology*, 24(18), R875–78.

Slovic, Paul, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor (2004), "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality," *Risk Analysis*, 24(2), 311–22.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein (1981), "Perceived Risk: Psychological

Factors and Social Implications," *Proceedings of the Royal Society A*, 376(1764), 17–34.

——— (1982), "Why Study Risk Perception?," *Risk Analysis*, 2(2), 83–93.

Slovic, Paul and Ellen Peters (2006), "Risk Perception and Affect," *Current Directions in Psychological Science*, 15(6), 322–25.

Sniezek, Janet A. and Timothy Buckley (1995), "Cueing and Cognitive Conflict in Judge-Advisor Decision Making," *Organizational Behavior and Human Decision Processes*, 62(2), 159–74.

Soll, Jack B. and Richard P. Larrick (2009), "Strategies for Revising Judgment: How (and How Well) People Use Others' Opinions," *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(3), 780–805.

Soon, C. S., A. H. He, S. Bode, and J.-D. Haynes (2013), "Predicting Free Choices for Abstract Intentions," *Proceedings of the National Academy of Sciences*, 110(15), 6217–22.

Sparrow, Betsy, Jenny Liu, and Daniel Wegner (2011), "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips," *Science*, 333(6043), 776–80.

Spiller, Stephen a, Gavan J Fitzsimons, John G Lynch, and Gary H McClelland (2013), "Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression," *Journal of Marketing Research*, 50(2), 277–88.

Tajfel, Henri and John C. Turner (1986), "The Social Identity Theory of Intergroup Behaviour," in *Psychology of Intergroup Relations*, ed. Stephen Worchel and William Auston, London: Burnham, 7–24.

Taylor, Matthew E. and Peter Stone (2007), "Cross-Domain Transfer for Reinforcement Learning," in *Proceedings of the 24th International Conference on Machine Learning - ICML '07*, New York, New York, USA: ACM Press, 879–86.

Taylor, Shirley and Peter A. Todd (1995), "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research*, 6(2), 144–76.

Thompson, Ronald L., Christopher A. Higgins, and Jane M. Howell (1991), "Personal Computing: Toward a Conceptual Model of Utilization," *MIS Quarterly*, 15(1), 125.

Turing, Alan M. (1950), "Computing Machinery and Intelligence," *Mind*, 59(236), 433.

Turpin, G. and T. Grandfield (2010), "Electrodermal Activity," in *Encyclopedia of Stress*, ed. George Fink, Amsterdam: Elsevier, 899–902.

Vázquez, Marynel, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson (2017), "Towards Robot Autonomy in Group Conversations: Understanding the Effects of Body Orientation and Gaze," *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, 42–52.

Venkatesh, Viswanath and Fred D Davis (2000), "A Theoretical Acceptance Extension Model: Four Longitudinal Field Studies," *Management Science*, 46(2), 186–204.

Venkatesh, Viswanath and Michael Morris (2000a), "Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force," *Personnel Psychology*, 53(2), 375–403.

——— (2000b), “Why Don’t Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior,” *MIS Quarterly*, 24(1), 115–39.

Venkatesh, Viswanath, Michael Morris, and Fred Davis (2003), “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly*, 27(3), 425–78.

Wang, Shensheng, Scott O Lilienfeld, and Philippe Rochat (2015), “The Uncanny Valley: Existence and Explanations,” *Review of General Psychology*, 19(4), 393–407.

Waytz, Adam, Kurt Gray, Nicholas Epley, and Daniel M. Wegner (2010), “Causes and Consequences of Mind Perception,” *Trends in Cognitive Sciences*, 14(8), 383–88.

Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang (2016), “A Survey of Transfer Learning,” *Journal of Big Data*, 3(1), 9.

Winton, Ward M., Lois E. Putnam, and Robert M. Krauss (1984), “Facial and Autonomic Manifestations of the Dimensional Structure of Emotion,” *Journal of Experimental Social Psychology*, 20(3), 195–216.

Yamada, Yuki, Takahiro Kawabe, and Keiko Ihaya (2013), “Categorization Difficulty Is Associated with Negative Evaluation in the ‘Uncanny Valley’ Phenomenon,” *Japanese Psychological Research*, 55(1), 20–32.

Yeomans, Michael, Anuj Shah, Sendhil Mullainathan, and Jon Kleinberg (2019), “Making Sense of Recommendations,” *Journal of Behavioral Decision Making*, forthcoming.

Yeomans, Mike (2015), “What Every Manager Should Know About Machine Learning,” *Harvard Business Review*.

Youyou, Wu, Michal Kosinski, and David Stillwell (2015), “Computer-Based Personality Judgments Are More Accurate than Those Made by Humans,” *Proceedings of the National Academy of Sciences*, 112(4), 1036–40.

Zhang, Baobao and Allan Dafoe (2019), *Artificial Intelligence: American Attitudes and Trends*, Future of Humanity Institute at Oxford University.

Zlotowski, Jakub, Diane Proudfoot, and Christoph Bartneck (2013), “More Human Than Human: Does The Uncanny Curve Really Matter?,” *Proceedings of the HRI 2013 Workshop on Design of Humanlikeness in HRI*, (January), 7–13.

APPENDIX A: STIMULI USED IN ESSAY 2

Study 1

Algorithm condition

Algorithms are a set of steps that a computer can use to accomplish a task. Thanks to rapid progress in computer science, algorithms can now be used to accomplish a wide range of tasks. Please use the sliders to indicate how much you would trust a computer algorithm to perform each of the tasks below.

Human condition

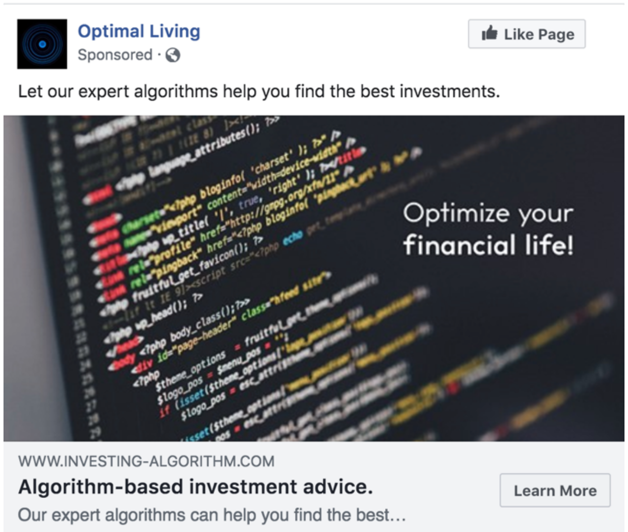
Please use the sliders to indicate how much you would trust a human to accomplish each of the tasks. For each task, consider a human who you think would be very well qualified for performing the task. For example, maybe a good friend would be well qualified for recommending a movie, but a doctor would be well qualified at diagnosing a disease. Please indicate how much you would trust this "well qualified human" for each task.

Study 2

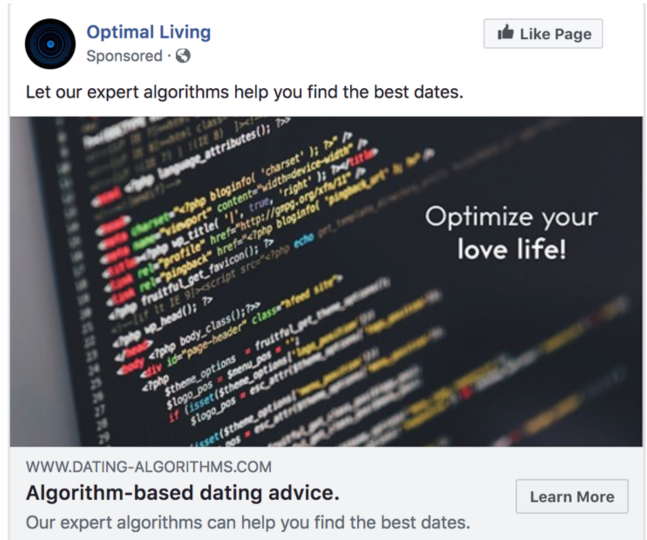
Human Dating Advice

Human Investment Advice

Algorithm Investment Advice



Algorithm Dating Advice



Study 3

[Performance data provided conditions included the paragraphs in italics about performance. Performance data not provided conditions omitted this information.]

Algorithms are a set of steps that a computer can use to accomplish a task. Thanks to rapid progress in computer science, algorithms can now be used to accomplish a wide range of tasks.

We are interested in whether you would trust algorithms more than a well-qualified human for several tasks listed below. For each task, please read the information describing the performance of an algorithm and the performance of a human, and then indicate which you would trust more for that task.

For each task, you can click on the "recent study" link to learn more. Each study was conducted by professional academic researchers and was reviewed by anonymous scientists to ensure that the research was conducted properly.

A recent study conducted by professional academic researchers showed that an algorithm can predict someone's personality based on their Facebook likes 14% more accurately than the person's own friends.

Who would you trust more to predict your personality?

A recent study conducted by professional academic researchers showed that cars driven by algorithms experience roughly 3.2 accidents per 100 million miles driven, compared to 4.1 accidents per 100 million miles driven by human drivers. Cars driven by algorithms can therefore be seen as about 28% safer than cars driven by humans.

Who would you trust more to drive a car?

A recent study conducted by professional academic researchers showed that an algorithm can predict how funny someone will find a joke with 61% accuracy, whereas the person's close friend can predict how funny they will find a joke with 57% accuracy. The algorithm was therefore about 7% more accurate than the person's own friend.

Who would you trust more to predict how funny you would find a joke?

A recent study conducted by professional academic researchers showed that an algorithm can recommend a treatment plan for cancer better than a human doctor. The algorithm was able to recommend the same treatment plan as human doctors in 990 out of 1000 cases, and also identified treatment options that the human doctors had missed in 300 of the cases. Some of these options were based on research papers that the human doctors had not read, since so much new research is constantly being published, but the algorithm was able to read all published research very quickly.

Who would you trust more to recommend a treatment plan for a cancer diagnosis?

A recent study conducted by professional academic researchers showed that an algorithm can predict what movies people will like with 20% more accuracy than human predictors.

Who would you trust more to recommend a movie?

A recent study conducted by professional academic researchers showed that an algorithm can diagnosis a psychological disorder 15% more accurately than an professional human psychologist.

Who would you trust more to diagnose a psychological disorder?

A recent study conducted by professional academic researchers showed that an algorithm can predict whether a criminal will re-offend during their parole with 86% higher accuracy than professional human judges.

Who would you trust more to decide whether a criminal is granted parole?

A recent study conducted by professional academic researchers showed that an algorithm can predict a college student's GPA with 107% higher accuracy than human admissions officers.

Who would you trust more to decide whether a student is admitted to college?

A recent study conducted by professional academic researchers showed that an algorithm can predict an employees job performance with 380% higher accuracy than a human interviewer.

Who would you trust more to decide whether an employee is hired?

Study 4

Subjective condition, movies

Science has shown that the kinds of movies people enjoy are based on their subjective moods and emotions, which means that knowing what movies someone enjoyed in the past is not always a good indicator what they will enjoy in the future. Predicting someone's enjoyment of movies is therefore a relatively subjective (vs. objective) task. Who would you trust more for the subjective task of recommending a movie: an algorithm, or your friend?

Subjective condition, dating

Studies have also shown that using subjective feelings and intuitions is the best way to choose who to date. Relying on intuition or gut feelings results in better romantic matches than relying on objective data like personality traits, likes, and dislikes. Recommending romantic partners is therefore a relatively subjective (vs. objective) task. Who would you trust more for the subjective task of recommending someone to go on a date with: an algorithm, or a professional match-maker?

Objective condition, movies

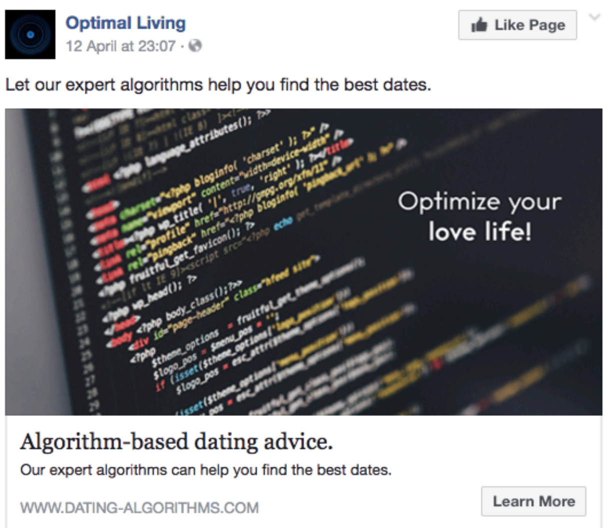
Science has shown that there are very clear patterns in what movies people enjoy, which means that knowing what movies someone has enjoyed in the past is a very good indicator of what they will enjoy in the future. Predicting people's enjoyment of movies is therefore a relatively objective (vs. subjective) task. Who would you trust more for the objective task of recommending a movie: an algorithm, or your friend?

Objective condition, dating

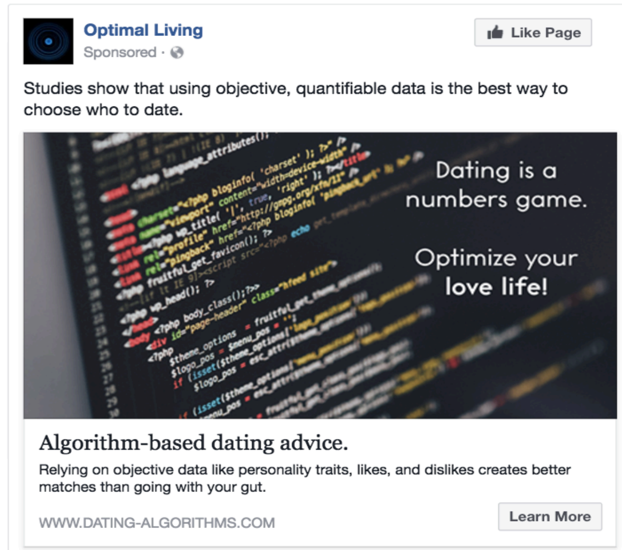
Studies have also shown that using objective, quantifiable data is the best way to choose who to date. Relying on objective data like personality traits, likes, and dislikes results in better romantic matches than relying on intuition or gut feelings. Recommending romantic partners is therefore a relatively objective (vs. subjective) task. Who would you trust more for the objective task of recommending someone to go on a date with: an algorithm, or a professional match-maker?

Study 5

Control condition



Objective condition



Study 6

High human-likeness condition

Artificial Intelligence (AI) is now capable of performing many tasks that only humans could do before. The best AI today can not only beat humans at chess and Jeopardy! – it can do a lot of things that were long thought to be fundamentally human:

-Compose music
-Create paintings
-Predict lasting romantic matches
-Predict which songs will be hits
-Write poetry
-Understand people's emotions

Research has shown that AI can do each of these things at least as well as humans can. Machines are therefore becoming much more human-like.

Low human-likeness condition

Artificial Intelligence (AI) is not capable of performing many tasks that only humans can do. The best AI today can not only beat humans at chess and Jeopardy! – but it can NOT do a lot of things that are thought to be fundamentally human:

-Compose music
-Create paintings

-Predict lasting romantic matches
-Predict which songs will be hits
-Write poetry
-Understand people's emotions

Research has shown that AI can't do any of these things at least as well as humans can. Machines are therefore not close to being human-like.

[All participants then read]:

In this survey, you will be asked to estimate the future value of the S&P 500, which represents the value of the 500 biggest companies listed on the New York Stock Exchange.

1 year ago today, the S&P 500 was listed at 2564 points.

Over the past year, the S&P 500 was listed at its highest point, 2930 points, on September 30, 2018.

Over the past year, the S&P 500 was listed at its lowest point, 2581 points, on February 8, 2018.

Below is a graph of the S&P 500's value over the past year.

Your task is to estimate the value of the S&P 500 one month from today.

Participants in this survey whose estimate is in the 5% most accurate estimates will be awarded a $3 bonus payment.

Please enter your estimate in the space below.

[After entering initial estimate, all participants read]:

A financial investment firm recently designed a computer algorithm that can estimate the future value of stocks. Its estimates are more accurate than the average estimates from recently graduated MBA students 80% of the time. This algorithm also estimated the value of the S&P 500 one month from now.

[Participants then read one of the following]:

Subjective task framing condition

Studies have shown that using subjective feelings and intuition is the best way to estimate the price of stocks, because feelings and intuitions are what drive many investors' decisions to buy or sell stocks. Relying on intuition or gut feelings therefore results in better estimates of a stock's future value than relying on objective data. Estimating the price of stocks is therefore a relatively subjective (vs. objective) task.

This algorithm can estimate the price of stocks well because it can model the effects of feelings and intuitions with high precision.

The algorithm estimated that the S&P 500 would be worth 2840 points one month from now.

Your estimate was ___ points.

In light of this new information, you can now change your original estimate or keep it the same. Please enter your estimate below.

Objective task framing condition

Studies have shown that using objective data such as supply and demand and other economic measures is the best way to estimate the price of stocks, because there are clear mathematical relationships between these economic measures and the future price of a stock. Relying on objective data therefore results in better estimates of a stock's future value than relying on subjective feelings or intuitions. Estimating the price of stocks is therefore a relatively objective (vs. subjective) task.

This algorithm can estimate the price of stocks well because it can model the effects of economic and historical data with high precision

The algorithm estimated that the S&P 500 would be worth 2840 points one month from now.

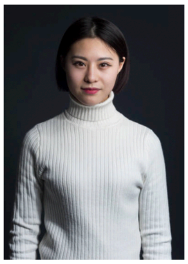
Your estimate was ___ points.

In light of this new information, you can now change your original estimate or keep it the same. Please enter your estimate below.

APPENDIX B: STIMULI USED IN ESSAY 3

Study 1

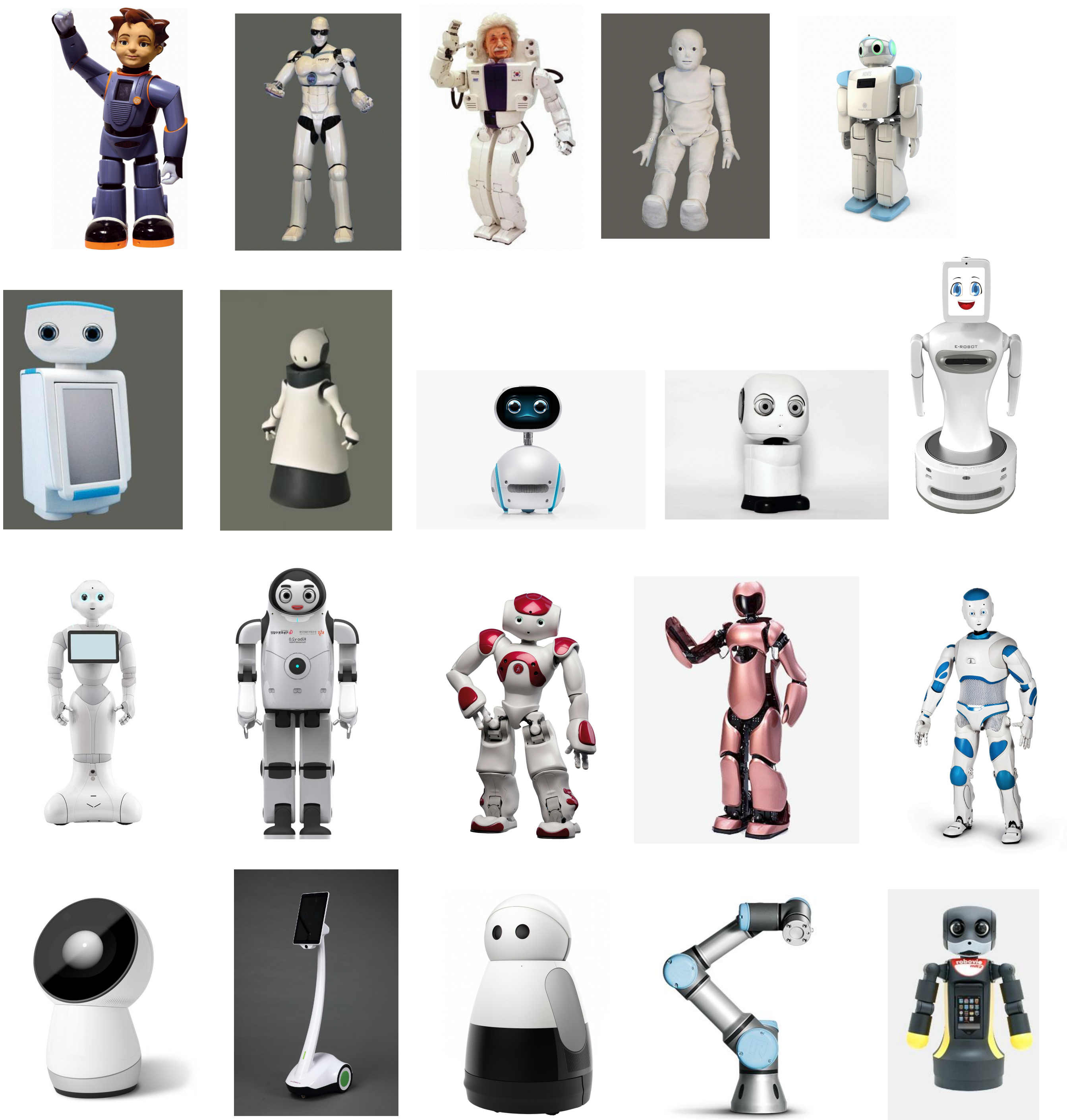
Humans (portrayed either as humans or as robots; participants saw one of four)




Robots (participants saw one from each of the five following rows)







After seeing each robot or human, participants answered the following questions, all on 0–10 scales:

How creeped out does this robot [person] make you feel?

How comfortable would you feel interacting with this robot [person]?

How human-like does this robot look overall?

Study 2

Participants watched one of the following videos:

Robots can have a mind: https://www.youtube.com/watch?v=SI4xaihN8Nk

Robots can't have a mind: https://www.youtube.com/watch?v=QvWQafW3NWg

Participants were then asked:

Please write a few sentences about why the arguments in the video are likely to be true.

Do you believe that robots can eventually have conscious minds?

Participants were then shown the following text and picture:

Several companies are developing robots that look and act almost perfectly like real humans. These robots are being used by businesses as receptionists, salespeople, waiters, and more, and by individual people as social and even romantic companions.

One of these robots, named Sophia, was even granted citizenship by Saudi Arabia earlier this year. The picture below shows one of these advanced humanoid robots. Please answer the following questions about this robot.

Participants then answered the following questions, on 0-10 scales unless specified:

How comfortable would you be shopping in a store where this kind of robot is employed?

How comfortable would you be dining in a restaurant where this kind of robot is employed?

Using the scale below, please indicate the position that best describes your overall evaluation of a company that would employ this kind of robot. [negative/positive, dislike/like, bad/good, on 1-7 scales]

As part of our research on this topic, we will be donating $1 for every participant who completes this survey to an organization that works on human-robot relations. We want to give you the opportunity to decide which organization we donate to on your behalf.

The first organization is called The American Society for the Prevention of Cruelty to Robots. They work to advance the development of human-like robots, which they think will be good for society.

The second organization is called The Center for the Study of Existential Risk. They work to prevent the development of human-like robots, which they think will be bad for society.

Which organization would you like us to donate $1 to on your behalf?

Do you believe that this kind of robot can have a mind?

How engaging was the video you saw at the beginning of this survey?

How convincing was the video you saw at the beginning of this survey?

How knowledgeable did the person in the video seem?

Study 3

All participants read the following introduction.

Robots today look and behave more and more like humans. This image shows a robot called Erica, who is currently working as a TV news anchor in Japan. Other similar robots work as receptionists in hotels and department stores, and one robot called Sophia has even been granted citizenship in Saudi Arabia.

Participants then read one of the following:

Experience only:

Beyond *looking* very human-like, these robots also have some components of a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *emotional computing*, which allows them to experience and express real emotions just like we humans do.

This is possible because when humans feel an emotion, the brain produces a pattern of electrical activity than can be replicated in a robot, letting it feel and express emotions just like we do.

However, these robots *cannot* think or act with autonomy like humans do – other than emotion, everything they do and say must be pre-programmed by a human. These robots therefore have some parts of a conscious mind, but not all.

Autonomy only:

Beyond *looking* very human-like, these robots also have some components of a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *cognitive* *computing*, which allows them to think and act with autonomy, without needing to be pre-programmed by a human.

This is possible because when humans have an independent thought or make an autonomous decision, the brain produces a pattern of electrical activity that can be replicated in a robot, allowing it to plan and act just like we do.

However, these robots *cannot* experience emotions like humans do. These robots therefore have some parts of a conscious mind, but not all.

Complete mind:

Beyond *looking* very human-like, these robots also have the capacity for a conscious mind in the same way that we humans do. Specifically, these robots use a technology called *cognitive computing*, which allows them to think and act with autonomy, without needing to be pre-programmed by a human. When humans have an independent thought or make an autonomous decision, the brain produces electrical and chemical activity that can be replicated in a robot, allowing it to plan and act just like we do.

In addition, these robots use a technology called *emotional computing*, which allows them to experience and express real human emotions. When humans feel an emotion, the brain produces a different pattern of electrical and chemical activity than can be replicated in a robot, allowing it to feel and express emotions just like we do.
These robots therefore have the capacity for all the same components of a conscious mind that we humans do.

Participants then answered the following questions, all on 0–10 scales:

Do you believe that robots will eventually be able to experience emotion like humans can?

Do you believe that robots will eventually have autonomy like humans do?

How comfortable would you feel as a patient in a hospital where this kind of robot is employed as a nurse?

How comfortable would you feel as a patient in a hospital where this kind of robot is employed as a hospital administrator?

This kind of robot seems competent.

This kind of robot seems useful.

Study 4

Participants first watched one of the two videos used in Study 2. Those in the “mind” condition then read the following:

As explained in the video, robots do have the capacity for a conscious mind in the same way that we humans do.

Specifically, robots use a technology called cognitive computing, which allows them to think and act with autonomy, without needing to be pre-programmed by a human. When humans have an independent thought or make an autonomous decision, the brain produces a pattern of electrical activity that can be replicated in a robot, allowing it to plan and act just like we do.

In addition, robots use a technology called emotional computing, which allows them to understand and express emotions. When humans feel an emotion, they often have specific facial expressions and tone of voice, which allows robots to accurately perceive human emotions and then respond appropriately.

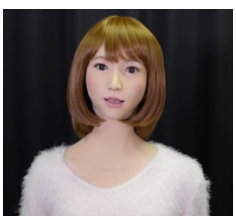
Robots therefore have the capacity for a mind just like we humans do.

Participants in the "no mind" condition read the following instead:

As explained in the video, robots cannot have a conscious mind in the same way that we humans do. We specifically mean that they cannot understand or express emotions, and they cannot think and act with autonomy – everything that they do, say, and experience must be pre-programmed.

Participants were then shown one of the two robots below and read:

Several companies are developing robots that are being used by businesses as receptionists, sales assistants, concierges, and more, and by individual people as social and even romantic companions.



Participants answered the following questions, all on 0–10 scales:

How much do you think this kind of robot could understand what you are thinking and feeling?

How much sympathy do you think this kind of robot would feel for you if you were suffering?

How competent do you think this kind of robot would be in real life?

How useful do you think this kind of robot would be in real life?

How much would you feel each of the following emotions during an interaction with this kind of robot? (uneasy, unnerved, creeped out)

Study 5

The mind perception manipulation was identical to Study 4.

Participants answered the following questions on 0–10 scales:

How much do you think this kind of robot could understand what you are thinking and feeling?

How much sympathy do you think this kind of robot would feel for you if you were suffering?

This kind of robot seems competent (agree/disagree).

This kind of robot seems useful (agree/disagree).