A Jagged Little Pill: Ethics, Behavior, and the AI-Data Nexus

Cameron F. Kormylo

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Information Technology

Idris Adjerid, Chair

Sheryl Ball

Paul Benjamin Lowry

Tabitha James

December 11, 2023

Blacksburg, Virginia

Keywords: Data, AI, Ethics, Mental health, Decision-Making, Privacy

A Jagged Little Pill: Ethics, Behavior, and the AI-Data Nexus

Cameron F. Kormylo

ABSTRACT

The proliferation of big data and the algorithms that utilize it have revolutionized the way in which individuals make decisions, interact, and live. This dissertation presents a structured analysis of behavioral ramifications of artificial intelligence (AI) and big data in contemporary society. It offers three distinct but interrelated explorations. The first chapter investigates consumer reactions to digital privacy risks under the General Data Protection Regulation (GDPR), an encompassing regulatory act in the European Union aimed at enhancing consumer privacy controls. This work highlights how consumer behavior varies substantially between high- and low-risk privacy settings. These findings challenge existing notions surrounding privacy control efficacy and suggest a more complex consumer risk assessment process. The second study shifts to an investigation of historical obstacles to consumer adherence to expert advice, specifically betrayal aversion, in financial contexts. Betrayal aversion, a well-studied phenomenon in economics literature, is defined as the strong dislike for the violation of trust norms implicit in a relationship between two parties. Through a complex simulation, it contrasts human and algorithmic financial advisors, revealing a significant decrease in betrayal aversion when human experts are replaced by algorithms. This shift indicates a transformative change in the dynamics of AI-mediated environments. The third chapter addresses nomophobia – the fear of being without one's mobile device – in the workplace, quantifying its stress-related effects and impacts on productivity. This investigation not only provides empirical evidence of nomophobia's real-world implications but also underscores the growing interdependence between technology and mental health. Overall, the dissertation integrates interdisciplinary theoretical frameworks and robust empirical methods to delineate the profound and often nuanced implications of the AI-data nexus on human behavior, underscoring the need for a deeper understanding of our relationship with evolving technological landscapes.

A Jagged Little Pill: Ethics, Behavior, and the AI-Data Nexus

Cameron F. Kormylo

GENERAL AUDIENCE ABSTRACT

The massive amounts of data collected online and the smart technologies that use this data often affect the way we make decisions, interact with others, and go about our daily lives. This dissertation explores that relationship, investigating how artificial intelligence (AI) and big data are changing behavior in today's society. In my first study, I examine how individuals respond to high and low risks of sharing their personal information online, specifically under the General Data Protection Regulation (GDPR), a new regulation meant to protect online privacy in the European Union. Surprisingly, the results show that changes enacted by GDPR, such as default choices that automatically select the more privacy-preserving choice, are more effective in settings in which the risk to one's privacy is low. This implies the process in which people decide when and with whom to share information online is more complex than previously thought. In my second study, I shift focus to examine how people follow advice from experts, especially in financial decision contexts. I look specifically at betrayal aversion, a common trend studied in economics, that highlights individuals' unwillingness to trust someone when they fear they might be betrayed. I examine if betrayal aversion changes when human experts are replaced by algorithms. Interestingly, individuals displayed no betrayal aversion when given a financial investment algorithm, showing that non-human experts may have certain benefits for consumers over their human counterparts. Finally, I study a modern phenomenon called 'nomophobia' – the fear of being without your mobile phone – and how it affects people at work. I find that this fear can significantly increase stress, especially as phone-battery level levels decrease. This leads to a reduction in productivity, highlighting how deeply technology is intertwined with our mental health. Overall, this work utilizes a mix of theories and detailed analyses to show the complex and often subtle ways AI and big data are influencing our actions and thoughts. It emphasizes the importance of understanding our relationship with technology as it continues to evolve rapidly.

ACKNOWLEDGMENTS

It is with profound gratitude that I write this section, acknowledging the invaluable support and guidance that have been instrumental in the completion of this doctoral dissertation. First and foremost, my deepest appreciation goes to Dr. Idris Adjerid, my dissertation chair, who for some reason believed in me enough to bring me out to Virginia as his first-ever PhD student. I will always be grateful for his patience through research funds being washed down the drain, panicked texts or strings of unrelated emails as I spill out any random thoughts I have, and, for the life of me, never being able to figure out how to write an academic contribution section. His expertise, mentorship, and friendship have been the cornerstone of this journey, providing an unwavering foundation upon which this work was built.

I extend heartfelt thanks to the rest of my committee members: Dr. Sheryl Ball, Dr. Paul Lowry, and Dr. Tabitha James. Thank you to Dr. Ball, whose passion and expertise in experimental economics not only enlightened my path but also kindled a deep-seated love for this methodology. Her dedication and insight have truly been a guiding light. Thank you to Dr. Lowry for his exceptional support and guidance in directing the Ph.D. program. His efforts in elevating the program to its current stature have not only benefited me but will continue to inspire future scholars. Thank you to Dr. James for the many lunches during which she imparted invaluable advice and guidance, enriching my understanding of what it means to be a scholar in our field. Her wisdom has been a treasured part of my academic journey.

I must also express my deepest thanks to my “unofficial” committee member, Dr. Corey Angst, for his unwavering support and belief in me. Through both his faith in offering me an opportunity to be a visiting PhD student at Notre Dame and his tireless work to turn that

opportunity into what is now a tenure-track position, his belief in me has never failed to help me believe in myself.

Most importantly, I have no words to express my heartful gratitude to my family. To my wife, Maddy, for her tolerance of my many late nights and endless stress along with her pride in my work that far surpasses my own pride, thank you will never be enough. To my children, Jamie and Lena, whose love and excitement every time I walk back into the house makes any hard day at work worth it, I have learned just as much from you in these last three years as you have from me. To my parents, thank you for your endless support and encouragement, for instilling in me the value of education, and for always pushing me to be the best version of myself.

Last but not least, to Alanis Morrissette, from whom the title of this dissertation draws inspiration, thank you for making music that can get anyone through the hardest nights of a doctoral degree.

This dissertation stands as a testament to all of your contributions, and it is with immense gratitude that I acknowledge your pivotal roles in this journey.

Contents

Introduction

Data, algorithms, and the rise of artificial intelligence have drastically shaped the landscape of today's society. Individuals are continuously being tasked with making complex decisions concerning their own digital privacy, when to interact with or avoid interacting with algorithmic decision makers, and how to balance their own well-being with the increasing grasp that these technological advances have on their lives. On one hand, the rise of big data and artificial intelligence (AI) has contributed to some of the most important advancements of the 21st Century. AI has served as a catalyst for clinical diagnostics, human connectivity, and economic growth. Looking forward, AI may even be a key driver in decelerating climate change or optimizing responses to future pandemics and disease outbreaks. Despite the important roles that AI serves, the consequences of such advances, left unchecked, may prove to be just as influential. The degradation of privacy rights has become apparent as data is increasingly collected and used for purposes far beyond consumers' comfort. Mental health continues to suffer as technological dependence is on the rise. AI has been trained on data that reflects the historical and systemic biases our society is attempting to distance itself from. To further complicate these issues, the black box nature of AI has given rise to serious issues of accountability as consumers struggle to assign blame when things go wrong.

AI undoubtedly has an important role to play in the coming decades. However, to ensure that it lives up to this potential, it is of increasing importance that it not remain unchecked. Central to this issue is that of *learning*. While considerable focus in methodological work considers how advanced machine learning methodologies allow a machine to learn, it is of similar importance to consider how humans learn to interact with the new world of data and AI. This dissertation contributes to that goal by examining the role of learning across several

contexts. First, I provide empirical insight into how consumers are able to navigate differing levels of privacy-related risk in light of enhanced privacy controls (Chapter 1). In that work I conduct a series of experiments that differ the presence of consent requirements laid out by the General Data Protection Regulation (GDPR) while differing inherent levels of privacy-risk. We find that in the high-risk conditions, consumers are influenced less by the presence or lack of privacy controls (implying that they recognize the high-risk nature of the decision and in either case move closer to the active choice level of consent). Contrarily, we find that low-risk privacy decisions are influenced to a greater degree by enhanced privacy controls. This implies that the impact of GDPR is most severe when the risk to consumer privacy is low.

Next, we examine the role of algorithms in moderating consumer decision biases (Chapter 2). Specifically, we examine the role of betrayal aversion, or the strong dislike for the violation of trust norms implicit in a relationship between two parties. Through an intricate financial market simulation, we are able to capture consumer preferences when faced with either a human financial advisor or an algorithm. We find that our participants are significantly impacted by betrayal aversion when faced with a human expert which decreased utilization by 16%. However, when the human expert was replaced with an algorithm, this effect was completely attenuated, and utilization of the algorithm was not impacted.

Finally, we explore the presence of nomophobia, or the modern fear of becoming disconnected from your mobile device, in the workplace (Chapter 3). Using a unique dataset collected from a large-scale field experiment, we show a strong effect of nomophobia on employees, operationalized as the stress response resulting from the decrease in mobile phone-battery level. We find several interesting moderating effects including demographics and personality. Additionally, we explore the impact of this phenomenon on workplace productivity.

Overall, this work provides important contributions to the broader literature on human behavior and data. I highlight the important link between the processing and collecting of data and the subsequent implications individuals have with AI and other advanced technologies. Making use of interdisciplinary theoretical constructs and robust methodological design, we highlight some of the important consequences that have arisen from this AI-data nexus.

Chapter 1 - Killing the Bees to Stop the Roaches: How the Homogeneity of Enhanced Privacy Protection Across Levels of Risk May Be an Overreaction

ABSTRACT

Protecting consumer privacy protections while maintaining a flourishing data economy has grown increasingly complex in recent years, particularly with the introduction of major regulatory changes such as the General Data Protection Regulation. One important, and often contentious, aspect of privacy protection is how to solicit consent from consumers. Policymakers and privacy advocates highlight the use of manipulative dark patterns utilized by firms that drive consent decisions concerningly high. Meanwhile industry advocates express concern that enhancing protections for data consent can result in an exodus of personal information from the data economy, decreasing innovation. We inform this debate by leveraging the political science literature on policy overreactions. We posit that while enhanced privacy consent may be essential, it can also be an example of a policy overreaction in lower risk settings. Through three behavioral experiments, we task participants with making privacy decisions in lower versus higher risk settings. The presence of dark patterns and enhanced privacy protections are manipulated across conditions. Our results show that protections such as protective choice defaults, may have similarly strong effects when privacy risk is lowered. Further, we find that introducing reversible consent may enhance the baseline effect of defaults, driving consent further down when paired with a protective default. Finally, we attempt to temper these effects by utilizing a social norm nudge but find that this elicits backlash from consumers. While recognizing the criticality of protecting consumer privacy, our work highlights the potential risk of new legislative protections becoming policy overreactions.

1. Introduction

A new wave of privacy regulation around the world has re-invigorated the debate over protecting consumer privacy while also encouraging the important innovations afforded to us through the data economy. For example, some experts have controversially claimed that the General Data Protection Regulation (GDPR), which became European Law in 2018, creates an illusion of privacy for a few at the expense of the many.[1] Similarly, the American Enterprise Institute presented their fears over GDPR to the Senate Judiciary Committee in 2019 citing that, among other issues, the law is cost-prohibitive for small and medium firms and that it threatens innovation and research (Layton 2019). A stream of active research substantiates some of these concerns and provides evidence that the introduction of stringent privacy regulation can have negative effects on the digital economy (Gal and Aviv 2020; Goldfarb and Tucker 2011; Janssen et al. 2022; Jia et al. 2021).

Privacy advocates instead highlight the dire need for enhanced protections, often citing the use of dark patterns, a term first introduced in 2010 by UX designer Harry Brignull[2], and the astronomical rates of online consent. For example, in 2020, the Federal Trade Commission (FTC) filed a complaint against an online webservice ABCmouse, for their use of dark patterns. In their complaint, the FTC define dark patterns as "design features used to deceive, steer, or manipulate users into behavior that is profitable for an online service, but often harmful to users or contrary to their intent" (FTC 2020). The complaint uses Brignull's term "roach motel" to describe a process that is easy to get in, but nearly impossible to get out. Examples of these roach motels could include free trials that automatically turn into paid subscriptions, repetitive email

[1] https://www.brookings.edu/blog/techtank/2018/06/11/a-case-against-the-general-data-protection-regulation/
[2] https://www.deceptive.design/

newsletters that clutter your inbox, or, particularly concerning, data consent decisions that are incredibly difficult to reverse. Other examples of dark patterns include "bad defaults" or as this paper will term them, dark defaults, such as universal opt-ins for consent choices (Cara 2019).

Concern over the status quo for privacy choice online has been a point of focus for policy makers. In the context of consumer consent and combating dark patterns, recent regulation, such as GDPR, often prohibits the use of dark defaults and increasingly requires that more protective defaults are used when soliciting consent. Further, recent regulation requires data permissions to be reversible and revisited on a regular basis. A growing body of literature studying the role of choice architecture (e.g., defaults) in privacy choice settings suggests that these changes required by new regulation will have a significant impact on consumer privacy choices (Acquisti et al. 2017; Adjerid et al. 2019; Egelman et al. 2013; Keller et al. 2011; Thaler and Sunstein 2008; Thaler et al. 2013).

However, whether these policy changes will have *only* their intended effect is unclear. For instance, Peer and Acquisti (2016) (the only work we are aware of evaluating reversibility in privacy settings) find that providing reversible privacy choices has the counter-intuitive effect of decreasing disclosure (despite objectively reducing the risk of disclosure). Equally concerning is whether consumers will be discerning in their response to more protective choice architecture (e.g., defaults). Ideally, these changes intended to protect consumers will alter consumer consent in settings where the risk to them is high but will have little effect when privacy risk is low. Alternatively, and problematically, a potential cost of these protections is that consumers could revoke consent in a blanket fashion, even when privacy risk is low. If this occurs, these policy changes can reduce the flow of personal information in ways that have minimal impact on privacy risk but potentially significant negative effects on firms and the data economy.

Critically, privacy protections afforded by these laws have been, in practice, broadly applied to data consent contexts, even when privacy risk may be low. Using GDPR as an example, if data collectors wish to bypass stringent consent requirements, they must establish a "legitimate interest", which ensures that there exists a compelling interest for the data collection that is not outweighed by the interests or rights of the data subject[3]. However, this is often a precarious path for many data collectors in light of the potentially significant fines if the legitimate interest is not upheld in court. Firms that have stated their desire to rely on a legitimate interest over obtaining consent have received legal threats and public backlash[4]. Additionally, a recent controversial decision by the Belgian Data Protection Authority stated that legitimate interest is not an adequate legal bases for online advertising[5]. This has led to businesses generally developing a legitimate interest as a fail-safe to obtaining consent, not a replacement (Butterworth 2018).

Conversely, using consent as a legal basis for data collection provides a clear and relatively unambiguous path to avoid legal ramifications and protect data subjects' rights. However, when using consent as the legal basis for data collection, the rigid requirements laid out by GDPR are required of any data-collectors, regardless of the risk to consumer privacy they impose[6]. Further, while GDPR and similar regulatory changes have only *explicitly* required the use of an active choice consent structure, not protective privacy defaults (e.g., the choice to *not* consent is pre-selected), cross-country variations in the interpretation of such requirements (Custers et al. 2018; Ruohonen and Hjerppe 2022) have resulted in mixed guidance as to the structure of the consent choice. For example, when subscribing to email newsletter, Ireland does

[3] https://gdpr-info.eu/art-6-gdpr/
[4] https://techcrunch.com/2022/07/12/tiktok-pauses-privacy-policy-switch/
[5] https://www.autoriteprotectiondonnees.be/publications/decision-quant-au-fond-n-21-2022-english.pdf
[6] https://gdpr-info.eu/art-7-gdpr/

not enforce requirements for explicit affirmative consent[7] whereas Germany requires a stringent double opt-in consent[8].

In settings with little to no risk for consumer privacy, the benefits of these requirements may be limited while the social costs associated with consumer behavioral changes may be high. We look to the political science literature and explore if this relationship between the costs and benefits of privacy regulation may indicate a *policy overreaction* where the policy "imposes objective and/or perceived social costs without producing offsetting objective and/or perceived benefits" (Maor 2012). To our knowledge, no work evaluating the impacts of recent regulatory changes to consent solicitation has considered the notion of a policy overreaction. Therefore, in this paper we ask: 1) are enhanced privacy protections effective at mitigating the effects of dark patterns used by firms (e.g., dark defaults, roach motels, etc.), 2) do consumers differentiate between levels of risk when navigating enhanced privacy protections (e.g., does consent significantly decrease even when privacy risk is lower?), and 3) can a behavioral nudge reduce the costs of a policy overreaction while allowing for the desired effect of privacy protections in higher risk settings?

To address our research objectives, we conduct three behavioral experiments with participants recruited from both Prolific and Amazon Mechanical Turk. In Experiments 1 and 2, participants are tasked with a decision that simulates logging-in to a webservice. In the high-risk setting (Experiment 1) the questions ask for sensitive disclosures such as, "Have you ever encouraged someone to drink in order to seduce them?" whereas in the low-risk setting (Experiment 2) logging-in may expose participant answers to questions such as "Do you prefer

[7] https://www.dataprotection.ie/en/organisations/rules-electronic-and-direct-marketing
[8] https://www.jdsupra.com/legalnews/germany-updates-privacy-guidance-6037413/

dogs over cats?" The login choice is differed across conditions by implementing two commonly used dark patterns and their respective enhanced privacy protection: 1) dark defaults (universally opting participants into the login decision) versus a privacy protective choice architecture (such as an active choice or a default opt-out) and 2) irreversible roach motels vs reversible consent.

In Experiment 3, we incorporate the two risk levels into one experimental design to allow for a direct comparison of treatment effects across levels of risk. Further, we increase the realism of the setting by presenting the study as a sign-up for a mobile application pilot. Participants believe that by partaking in the study, they will have to download an application on their personal device and use the app for four weeks. Risk levels are differentiated across conditions by manipulating the mobile permissions asked for by the app (either trivial or intrusive). Participants are then asked to consent to the researchers sharing their app data with third party corporate sponsors to aide in the development of future applications. This choice is also differentiated by default structure and reversibility.

Finally, we introduce a social norm nudge that tells users that 60% of past participants chose to consent to sharing their data with third parties. These nudges are attractive due to an individual's inclination to use the behavior of their peers as a suitable guide for their own behavior (Nahmias et al. 2019) and their potential to address the consequences of a policy overreaction by signaling to users that a particular setting is lower risk and thus encouraging them to continue to provide consent. We explore the interaction of our two dark patterns (dark defaults and roach motels) with this social nudge to evaluate whether any costs of a policy overreaction we observe in Experiment 1 and 2 can be tempered.

Results from our first two experiments provide several interesting findings. First, we confirm that enhanced privacy controls are highly effective at decreasing consent rates in higher

risk settings. However, and problematically, these effects persist in lower risk settings. In addition, we find that introducing reversibility enhances further entrenches the effect of protective defaults. Again, these effects are insensitive to disclosure risk level. For example, in the lower-risk setting adding reversibility to the protective default decreases login rates by an *additional* 13.8%. These findings highlight both the potential of these protections to reduce consent in higher risk settings, but also, their potential to have problematic spillover effects to lower risk settings.

Experiment 3 provides enhances the realism of our experimental context and manipulates risk levels simultaneously in a single experiment. The results of experiment 3 substantiate the prior findings. First, we find that attempts to elicit higher levels of opt-in through dark default had less efficacy in higher-risk setting, demonstrating that participants are sensitive to increases in privacy risk. However, protective defaults continued to decrease consent rates for both high and low risk disclosure settings. Finally, we find that the social norm nudge was not able to temper the strong decline of consent in lower-risk settings. Instead, this nudge elicited a backlash from decision makers in higher-risk settings, suggesting that users viewed it as an attempt to manipulate their consent levels. These results highlight the potential consequences of modern privacy regulation and the difficulty of reversing those unintended costs, lending support to the idea that there may be a policy overreaction.

This work extends a budding literature on the impact of privacy regulation on various firm and consumer outcomes (Adjerid et al. 2015; Goldfarb and Tucker 2011; Miller and Tucker 2017), particularly the newer streams of work evaluating the most recent wave of global privacy regulation (Godinho de Matos and Adjerid 2021; Goldberg et al. 2019; Janssen et al. 2022). Leveraging the lens of policy overreaction from political science, we examine the often posited

but, to our knowledge, scarcely investigated conjecture that enhanced privacy protection (i.e., enhanced consent protections) can result in an undue decline of data allowances online. Specifically, we show that even in settings where individuals face little risk to their privacy, the effects of enhanced privacy controls, such as protective defaults and reversible consent, may be just as strong in lower vs. higher risk settings. In addition, we differ from most recent work examining the impact of regulation in that we use an experimental approach (most prior work has focused on field evaluations of these regulations). Our approach, although more removed from the regulation that inspires it, allows us the to break down effects of each privacy protection and more precisely identify sources of policy overreaction. In addition, this approach allows us to evaluate how different aspects of privacy protections interact with each other. For example, our results imply that reversible consent, while having little effect in isolation, has a strong interaction effect that enhances the directional effects of choice defaults.

The second stream of work we contribute to is the literature on privacy decision making and the role of behavioral biases, dark patterns, and choice architecture in how consumers make privacy choices (Acquisti et al. 2017; Adjerid et al. 2019; Egelman et al. 2013; Johnson et al. 2012; Peer and Acquisti 2016; Thaler et al. 2013; Waldman 2020). Notably, most of this literature has evaluated how behavioral biases contribute to inconsistencies in privacy decision making, such as inexplicably high levels of consent (Acquisti and Grossklags 2005), and how choice architecture can help alleviate some of these issues (Thaler et al. 2013). The evolution of recent privacy regulation to explicitly define the choice architecture around privacy protection seems to reflect the conclusions of this stream of work. We extend this body of work by considering whether the wide proliferation of protective choice architectures, while reducing disclosure in sensitive settings (perhaps in line with the intention of these protections), can also

have an unintended effect of reducing disclosure of less sensitive data. This effect on less sensitive data likely has minimal effects on consumer privacy but can have significant effect on data availability for firms. In addition, we examine how various aspects of choice architecture can interact to exacerbate this effect (e.g., reversibility and protective defaults). Finally, we contribute to the emerging literature on dark patterns in privacy settings and show that attempts to nudge consumers into higher levels of consent can backfire when dark patterns are present.

2. Conceptual Background and Hypotheses

One important focus of modern-day privacy regulation is the proliferation of firm approaches, such as dark patterns, that take advantage of consumer limitations in decision making to encourage high rates of data sharing. Our work first asks if the privacy protections afforded by these regulatory changes are effective at combating those tactics. Therefore, we explore prior literature on dark patterns and their connection to privacy decision making and choice architecture – the idea that subtle changes in the design of a choice can drastically impact the behavior and decision making of an individual presented with that choice (Egelman et al. 2013; Johnson et al. 2012; Keller et al. 2011; Thaler et al. 2013). Next, we evaluate how the lack of regulatory differentiation between levels of privacy risk, as it relates to requirements for collecting consent, may elicit significant consumer and firm costs without offsetting privacy benefits. To do this, we first examine the current literature that addresses both the intended and unintended consequences of privacy regulation (Adjerid et al. 2015; Aridor et al. 2021; Godinho de Matos and Adjerid 2021; Goldberg et al. 2019; Goldfarb and Tucker 2011) before introducing literature from political science that focuses on policy overreactions (Maor 2012). Finally, we introduce social norm nudges as one potential solution to temper some of the costs that come from these regulatory changes.

2.1 Dark Patterns in Privacy Decision Making

The traditional view within classical economics has been that decision makers operate through a rational process, driven by the desire to maximize utility (Mullainathan and Thaler 2000). When applied to privacy decision making, this gave rise to the development of a privacy calculus, in which consumers systematically weighed the costs and benefits of privacy-related disclosures (Dinev and Hart 2006). However, the information systems literature is increasingly adopting insights and rationale from behavioral economics (Goes 2013). This adoption has led to a softening of these assumptions and posits that privacy choices are driven by both deliberative assessments of benefits and risks as well as by bounded rationality and cognitive biases (Acquisti et al. 2012; Li et al. 2008; Tsai et al. 2011).

Manipulations in the presentation of a choice, termed choice architecture, take advantage of these biases to encourage a particular decision outcome. The powerful effects of choice architecture have been highlighted extensively in the literature (Egelman et al. 2013; Johnson et al. 2012; Keller et al. 2011; Thaler et al. 2013).While choice architecture can be used to encourage positive behaviors, such as improving one's health (Quigley 2013), it can also be used by firms to manipulate individuals into sharing data or purchasing products (Brignull 2022). These uses of choice architecture have been termed *dark patterns*. Dark patterns are "user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that, if fully informed and capable of selecting alternatives, they might not make" (Mathur et al. 2019). The term was first coined by UX designer Harry Brignull in 2010[9]

[9] Birgnull maintains a website that documents the use of dark patterns. See: deceptive.design (formerly darkpatterns.org).

but has sense been adopted by policymakers, industry professionals, and academics to describe this widely experienced phenomenon.

While the study of dark patterns by researchers is still relatively new, there is some work that considers the role that dark patterns play in users' online decision-making. Gray et al. (2018)) show that dark patterns undermine a user's action possibilities, and in some cases even remove user choice entirely. Using the context of proxemics, Greenberg et al. (2014)) describe numerous types of dark patterns that use an individual's physical location to manipulate them to some desired action. For example, using a bait-and-switch strategy, one bus stop display developed technology to identify when an individual was looking directly at it. When individuals were not looking at it, it would display an eye-catching image in an attempt to attract viewers and then after they looked in its direction, the display would change to the intended target material.

Within the realm of privacy, dark patterns have been proposed as an explanation to the privacy paradox – the idea that despite user's stated privacy concerns, they fail to take reasonable action to protect their privacy (Waldman 2020). Further, privacy dark patterns, such as universal opt-in defaults, or dark defaults, have been shown to activate an individual's System 1 thinking to encourage fast, automatic, and often irrational decisions with the hopes of collecting more personal data (Bösch et al. 2016). Additionally, Baroni et al. (2021)) use the example of a roach motel dark pattern – a "trap" that is easy for users to get into but nearly impossible for them to get out (e.g., making it exceedingly difficult to end a paid subscription) to highlight the emotional and economic distress that dark patterns impart on consumers. Interestingly, some work has shown that while dark patterns are effective at driving consumer consent, more

aggressive dark patterns may elicit a downstream emotional backlash from the users, leading to a potential decrease in their future effectiveness (Luguri and Strahilevitz 2021).

In light of these findings, our first aim to explore the role of privacy regulation in altering the state of dark patterns. Specifically, we explore two dark patterns: choice defaults and reversibility of privacy choices (e.g., roach motel). Bothe of these patterns are powerful tools of choice architecture and central consideration of GDPR and similar regulations.

2.1.1 Choice Defaults

As defined by the choice architecture literature, a default is the first considered option when making a decision and is the "status quo" for the decision maker before they consider other options (Huh et al. 2014). Generally, defaults are presented with one option as the opt-out choice (meaning that the decision maker must exert effort to change their decision from that choice; e.g. the default) and all others as opt-in (Johnson and Goldstein 2003). Decision makers presented with these defaults are given a default configuration that they can then add or subtract features from (Park et al. 2000). Defaults could also present as a choice between two options but where only one is given and the other must be requested (McKenzie et al. 2006).

Defaults can either encourage positive behaviors, such as default amounts for charitable contributions (Goswami and Urminsky 2016) or defaults for optimal savings rates for 401(k) enrollment (what we are terming a "protective default") (Choi et al. 2003), or it could act as a dark pattern (what we are terming a "dark default") that opts users in to consenting to sensitive data sharing (Bösch et al. 2016). Whichever way a default is presented, extensive evidence exists showing that this specific choice architecture will have significant effects on consumer outcomes (Johnson et al. 2002; Johnson and Goldstein 2003; Thaler and Sunstein 2008). A classic example in the literature shows how presenting an organ donation decision with a default opt-in design

can drastically alter the decision maker's willingness to donate (Johnson and Goldstein 2003). Similar effects have been shown for consumer product choices (Brown and Krishna 2004; Dinner et al. 2011).

Discussions on the causes of default effects have provided several valuable findings. First, a default choice is often "easy." Identifying the best option among choices and analyzing underlying tradeoffs takes time and increases cognitive effort (Tversky and Kahneman 1974), whereas no effort is required of the decision maker when a default choice is presented. A decision maker exhibiting cognitive laziness may be more susceptible to a less effortful choice (Fiske and Taylor 1991; Samuelson and Zeckhauser 1988; Thaler and Sunstein 2008). Consumers have shown this affinity for less effortful choices when alternatives require higher levels of cognitive effort (Brown and Krishna 2004; Camerer et al. 2003; Johnson et al. 2002). Additionally, when decision makers are tired (Levav et al. 2010) or when their self-control has been taxed (Evans et al. 2011) they have been shown to be more susceptible to defaults. These effort based accounts show that default effects are most impactful when people fail to align the effort required to make the choice with the importance of its outcome (McKenzie et al. 2006). Given these extensive findings, we hypothesize that:

H1: *Protective choice architectures will decrease rates of consent.*

2.1.2 Reversibility

In addition to the changing structure of choice defaults, allowing for reversible consent is a main concern of most modern privacy regulation, most notably GDPR (Article 29 2017). These changes are intended to grant further control to consumers over their privacy and combat the irreversible nature of roach motel dark patterns. However, the true effects of these changes are

uncertain ex ante. Some work has shown that giving consumers control over their personal data is a major driver towards privacy related trust (Whitley 2009). When trust in an online entity is high, individuals tend to make riskier privacy decisions (Lauer and Deng 2007). This implies that allowing for reversible consent may make consumers more willing to disclose of personal information online.

However, findings that do substantiate reversibility as a trust-invoking element of control within online privacy are limited. By contrast, more recent work primarily shows the reverse. Peer and Acquisti (2016)) show that reversibility instead cues the individual as to the seriousness of the decision at hand. This leads to less disclosure, even if the risk is lowered as well. Additionally, they show that both reversibility and irreversibility have the same directional effect, so long as they are made salient. This implies that if a user is aware that they may be interacting with a roach motel, they may respond by decreasing their willingness to share. On the flip side, if a user is explicitly aware that the choice is reversible, this may cue the user as to the seriousness of the choice and similarly increase caution. In other words, while individuals may still prefer a reversible decision, this does not entice them to disclose more as it entices them to, say, purchase more from a vendor (Wood 2001). Therefore, we hypothesize that:

H2: *Making a consent decision explicitly reversible will lower rates of consent.*

2.2 The Intended and Unintended Effects of Privacy Regulation

There is a growing body of work evaluating the impact of various privacy regulations on diverse real-world outcomes (Goldfarb and Tucker 2011; Johnson et al. 2022; Miller and Tucker 2017). The results from this literature highlight a complex set of tradeoffs that impact both the profit potential of firms and consumer privacy and welfare.

Notable work has highlighted the increase in consumer privacy protection afforded by GDPR. For example, data from a large telecommunications provider in Europe showed an increase in opt-in rates post-GDPR for some allowances but found that more sensitive permissions were generally restricted (Godinho de Matos and Adjerid 2021). The same paper provided evidence that as a result of enhanced consent, customers who chose to opt-in drove higher sales and increased contractual lock-ins, a benefit for the firm. Additionally, an examination of privacy policies post-GDPR showed that the level of protection for personally identifiable information for children and data aggregation (two major privacy concerns) increased by 22% and 13% respectively (Zaeem and Barber 2020).

However, the economic impacts of these regulations are cause for concern. The primary unintended consequence of privacy regulation concerns the restriction of data sharing that prevents synergies and limits data-driven knowledge and innovation (Gal and Aviv 2020). This is highlighted by examining investments in technology ventures for GDPR exposed firms. Jia et al. (2021)) estimates a roughly $1.2 billion decrease in raised technology ventures for new firms in the EU after GDPR, extrapolating to between 3,604 and 29,819 lost jobs. Further highlighting the innovation costs of privacy regulation, Janssen et al. (2022)) find that after GDPR took effect, approximately one-third of available apps in the Google Play Store exited the market. Additionally, they find that consumer surplus and aggregate app usage declined by one third post-GDPR. Bleier et al. (2020)) also suggests that startup firms are likely to take their talents away from a post-GDPR Europe, restricting innovation and generally damaging the EU economy.

2.2.1 Policy Overreaction

This work posits that major regulatory changes like GDPR are important steps forward in protecting consumer privacy. However, we also look to the political science literature to explore if, in light of no risk-differential for consent requirements, these laws may constitute a *policy overreaction.*

A policy overreaction is a policy that "imposes objective and/or perceived social costs without producing offsetting objective and/or perceived benefits" (Maor 2012). One example is the regulatory requirement to show graphic warning labels on cigarette packages to stop individuals from smoking. While some evidence shows that the labels are effective at stopping new users from picking up the habit, research also shows that the overly graphic nature of the labels leads those with a preexisting nicotine addiction to experience a stress response that leads them to smoke *more* (Erceg-Hurn and Steed 2011). Policy overreactions, while potentially solving the initial concern, often bring about new, unintended problems as they may put significant focus on the effectiveness of the policy with little to no consideration of cost (Maor 2021).

Especially relevant to our consideration of privacy risk settings, Maor (2019)) describes one type of policy overreaction that occurs when heterogeneity around the policy context or the intended target of the policy are not considered. Examples include zero-tolerance policies or unlimited policies (e.g., closed border policies versus open border policies). In the following paragraphs, we explore the possibility of stringent consent requirements being a policy overreaction. Specifically, by failing to incorporate heterogeneity in policy requirements around levels of risk, the social cost of the policy is heightened. Given that in low-risk settings, the changes are unlikely to provide significant benefit to the population, the costs may not be outweighed. It is also important to state that an overreaction does not mean that *no reaction* is

warranted, our claim in the following paragraphs is that while enhanced privacy protections are important in many settings, the breadth of contexts to which they apply may lead consumers to prioritize privacy even when privacy is not at risk.

We focus specifically on the social costs associated with these regulatory changes. If consumers are scrupulous to levels of risk when faced with consent decisions, the lack of heterogeneity around consent requirements may not a pressing issue. However, if, instead, consumers have a blanket approach to consent that is heavily influenced by regulatory changes to choice architecture across any level of risk, then the cost of such policy is increased.

If we take protective choice defaults as the regulatory response to some specific privacy concern, we can begin to see how significant costs might arise. An individual with a pre-existing proclivity towards default effects (e.g., Dinner et al. (2011))) that additionally exhibits privacy-related anxieties and an action bias towards precaution (Rottenstreich and Hsee 2001), likely has a high chance of accepting a protective default. If the context is particularly risky from a privacy perspective, then this proclivity may be acceptable. However, if the decision involves little-to-no privacy-related risk, then the reaction likely *costs* more than it *benefits* (i.e., given the low-risk privacy setting, the decision-maker gains little in not consenting but may lose out on efficiency, reliability, personalization, etc.), implying that by not considering heterogeneity around risk, the policy becomes an overreaction.

Evidence from prior literature highlights the difficulty consumers face when contextualizing risk. For example, heightened emotion (such as that which follows a trigger event described above) can lead to the neglecting of risk probabilities and an often harmful overreaction to some threat (Sunstein and Zeckhauser 2011). One example used in the literature describes the effect of extensive security precautions at airports. These precautions elicit

anxieties in individuals leading to higher rates of driving over flying. Given that flying is objectively far safer than driving, the security precautions may actually lead to a loss of life (Sunstein and Zeckhauser 2011). Experimental economics has also highlighted this tendency in individuals. One study looked at individual responses to risk avoidance, differing the level of emotion elicited in their participants. The results showed that even when the probability of a risk is low, participants exhibited action bias and took precautionary steps even when those steps were not plausibly justified (Rottenstreich and Hsee 2001).

This idea is also reflected in the popular press. In one Forbes article[10], the author states, "A big reason for the jump in privacy concerns is primarily a result of consumers becoming more aware of how companies are using their data… With news stories breaking like the Cambridge Analytica Scandal…it's hard for consumers to ignore the importance of protecting their data." Similarly, an article reporting on the birth of GDPR[11] claims, "After months of learning about data breaches from companies like Facebook and Equifax, this couldn't be more necessary." These trigger events, as described by Birkland (2006)), seem to have created a shift in public opinion over data privacy leading to heightened anxieties and a potential for increased caution in online settings even when privacy is not at risk.

Given the extensive findings described above, we hypothesize:

H3: *Protective choice architectures will decrease rates of consent in both high and low risk settings.*

[10] https://www.forbes.com/sites/forbestechcouncil/2020/12/14/the-rising-concern-around-consumer-data-and-privacy/?sh=57557bf3487e
[11] https://www.forbes.com/sites/andrewrossow/2018/05/25/the-birth-of-gdpr-what-is-it-and-what-you-need-to-know/?sh=4852abc855e5

2.3 Social Norm Nudges

If certain regulatory changes do elicit heightened social costs that outweigh the perceived benefits in certain risk settings, one way to tamper these unintended effects would be to encourage a more thoughtful consumer approach to privacy risk level while making consent decisions. Nudging has consistently been shown to provide effective behavioral change in decision makers with little cost (Benartzi et al. 2017). One highly effective method commonly used to achieve this behavior change is to utilize social norm nudging. This method motivates decision maker behavior by providing a benchmark as to what decision is the most effective in a given context (Cialdini et al. 1991). Individuals have been shown to increase charitable contributions when informed of others' contributions (Frey and Meier 2004), decrease energy consumption when provided with neighborhood averages (Schultz et al. 2007), and be more likely to vote if they anticipate a high voter turnout (Gerber and Rogers 2009).

However, the effectiveness of social norm nudging in the literature is heterogeneous. For example, some work has shown that this method has little to no effect when attempting to increase tax compliance (Blumenthal et al. 2001) and in some cases, may actually *reduce* the rate of payments (John 2018). In the context of financial risk, some work has shown that social norm nudging has no effect in a high-stakes monetary settings (Mol et al. 2021). Even further, informing employees who had a 0% contribution rate to their 401(k) of their peers' savings rates made them less likely to increase their contribution (Beshears et al. 2015); an oppositional reaction that again highlights the heterogeneity of social information around the stakes of the decision context. Similarly, a large-scale field experiment showed that individuals who qualified for an earned income credit were *less* likely to claim it if they were informed that 4 out of 5 similarly situated peers had done so (Bhargava and Manoli 2015).

Interestingly, within the realm of privacy decision-making, past work has shown both positive and negative effects of social norm nudging on subsequent behavior. Using these nudges has been associated with a more privacy-focused approach to online cookies (Goecks and Mynatt 2005), a higher adoption of personal firewalls (Goecks et al. 2009), and, to a limited extent, enhanced privacy settings on messaging apps (Patil et al. 2011). Conversely, Acquisti et al. (2012)) show that individuals may increase privacy disclosures when exposed to a social norm nudge, but importantly are impacted significantly less when the disclosures are presented in a decreasing order of intrusiveness (priming participants to view the disclosure task as particularly sensitive), while the nudges have a strong effect when less sensitive questions are presented first (signaling a lower-risk setting).

Given these findings, we propose using a social norm nudge to temper the unintended effects of privacy regulation that decrease disclosure even when privacy is not at risk. The findings from past work suggest that in low-risk settings, users should be more influenced by social information than in high-risk settings. This would mitigate the loss of data by encouraging disclosure when privacy is not at risk but would avoid impeding on regulatory benefits by having a lower effect when privacy is a greater concern. Therefore, we hypothesize:

H4: *Informing individuals of past consent rates will decrease the effect of protective defaults in lower-risk settings but not in higher-risk settings.*

3. Experiment 1 & 2

Experiments 1 and 2 utilize similar experimental designs with a few notable exceptions described below (primarily the sensitivity of disclosures). Participants for these experiments were pulled from the same sample pool but were conducted at separate times. We present the

analyses from both experiments side-by-side to allow for a broad understanding of the trends measured for both levels of risk. In a third experiment, we test our hypotheses in a more realistic context and manipulate risk level in a single experiment simultaneously. We also introduce the social nudge in Experiment 3. The additions to Experiment 3 allow us to make direct comparisons of effects across risk level and evaluate whether the nudge impacts individual behavior.

3.1 Experiment 1

We utilized a two factor between-subjects online experiment that differentiates 1) the default structure of the consent choice and 2) the reversibility of the consent choice. We use the context of logging in to a webservice to investigate how enhanced privacy protections may impact this widely experienced task.

3.1.1 Default Structure

Participants were randomly assigned to one of three default structure conditions: default logged-in (our control dimension, indicative of the traditional universal opt-in structure or a "dark default"), active choice (the structure required by privacy law such as GDPR), and default not logged-in (a highly protective opt-out privacy structure encouraged, but not often required, by privacy regulation). Participants in the dark default and protective default conditions were presented with one choice for the login decision: "Sign into my research profile" or "Continue as guest" respectively. In either case, the option was pre-selected. If the participant wanted to change the selection, they had to manually uncheck the box. In the active choice dimension, they were presented with both of the above options: "Sign into my research profile" and "Continue as guest." Neither option was pre-selected, requiring these participants to manually choose one of the two options. The three default presentations can be seen in Appendix A.

3.1.2 Reversibility

We also manipulated the reversibility of the login decisions. Each participant saw their login decision presented with one of three options: a statement making the reversibility of the login decision explicit, a statement making the irreversibility of the login decision explicit, or no statement regarding reversibility (our control dimension). The reversible statement read, "This decision can be changed by contacting survey administrators." Conversely, the irreversible statement informed the participant that, "This decision cannot be changed and is final." Examples of these statements can be seen in Appendix A.

3.1.3 Repeatability

Given the repeated nature of consent elicitation brought about by modern privacy legislation (e.g., repetitive cookie warnings), Experiment 1 intended to measure the change in choice architecture effects over repetitions of the same choice. Three surveys were given to each participant where the decision to login was made at the onset of each.

The experiment was designed to mimic a user's repeated interactions with an online webservice. Measures were taken to ensure that the participants felt that each study was separate and unique. For example, at the end of each study, participants were asked to watch a video relating to the context of the specific study that acted as a time lag between studies. Interestingly, we found that the effects of our changes to choice architecture did not vary significantly over time (See Appendix B). Therefore, we chose to not include repeatability in subsequent experiments and analyze results from only the first of the three surveys shown to each participant.

3.1.4 Risk Level

Experiment 1 was designed to be our "high-risk" setting. In this task, participants answered questions that may or may not be linked back to them depending on their choice to login. These questions related to criminal history, sexual behavior, and romantic involvement. Examples include, "Have you ever had sexual thoughts about a member of your same sex?" and "Have you ever encouraged someone to drink when you were trying to seduce them?" The full set of questions can be found in Appendix C.

3.1.5 Participants and Procedure

1,600 participants were recruited from Prolific for Experiment 1. Participants were told that they would be answering questions related to unethical behavior that may be intrusive. After filtering out unengaged responses (abnormally quick completion times or failing attention checks) and those who asked to be removed from the study after their debrief, we were left with a sample size of 1,526. Participants were compensated $3.40 upon completing all three studies.

Participants began by creating a research profile. This process asked participants to report demographic information such as gender, ethnicity, geographical location, and education/work status. This research profile simulated a personal profile that one would create on a popular web service. Participants were then directed to take three studies, in a random order. Participants were informed at the beginning of each study that we were exploring potentially unethical behavior and that some of the questions that would be asked dealt with adult or sensitive material. The participants were also informed that they may skip any questions that they wish not to answer. The three studies contained questions related to criminal activity, sexual activity, and romantic involvement, respectively. These studies were randomly ordered and counterbalanced.

At the onset of each study the participants were given the choice to login to their research profile. This choice was structured dependent on their randomly assigned default and reversibility conditions. They were informed that logging in would allow us to track their responses and that by doing so, their responses would be linked back to them. To make the decision salient, participants who chose to sign into their research profile saw their Prolific ID listed at the top of each subsequent page of the study. Participants who chose to not login saw "Guest" listed instead.

The participant then was directed to a series of 5 questions related to the study topic. After the 5 questions were answered, they were shown a short video that they had to watch and subsequently summarize and reflect on. They were then randomly directed to one of the remaining studies. At the conclusion of the third and final study, they were asked a set of exit questions (see Appendix D) and were then debriefed.

3.2 Experiment 2

Our second experiment largely followed the same design used in Experiment 1. However, after completing the research profile, Experiment 2 participants were required to answer trivial, lower-risk questions. The question set included inquiries such as, "Do you prefer dogs over cats?" and "Do you like reading fiction more than non-fiction?" These questions pose little-to-no risk to the participants if their answers were to be identified back to them. The full set of questions can be found in Appendix C. Additionally, as stated above, Experiment 2 only presented participants with one study, not three.

500 participants were recruited from Prolific to take part in Experiment 2. After filtering out participants who gave unengaged responses and those who asked to have their data removed

after being debriefed, we were left with a sample size of 440. Participants were compensated $0.90 for completing the study.

3.3 Results

We evaluate observable variables for each participant (age, race, gender, education, and employment) and find balance across most variables in both experiments, with few exceptions. Descriptive statistics and a table of pairwise comparisons can be found in Appendix E. Our results from the first two experiments show that first, even in lower-risk settings default structures have large and significant effects on participant behavior. Second, we find a powerful interaction effect between default structures and reversibility in which both higher- and lower-risk settings see reversibility *increasing* the baseline default effect. In other words, when reversibility is paired with a dark default, login rates *rise* even further than they do when presented with a dark default alone. Conversely, pairing reversibility with a protective default generates a significant *drop* in login rates beyond the initial effect of the protective default.

3.3.2 Effect of Defaults

The dependent variable in our regression is a binary indicator variable that represents whether or not the participant *i* chose to login to their research profile. We estimate this model for both Experiments 1 and 2. For Experiment 1, we limit our analysis to their first exposure to the choice. The model used to estimate our results can be seen below.

$$Login_i = \beta_1 ActiveChoice_i + \beta_2 ProtectiveDefault_i + u_i$$

$ActiveChoice_i$ is a binary indicator for whether or not a participant was in the active choice condition. $ProtectiveDefault_i$ is a binary indicator for whether or not a participant was

in the protective default condition. Estimates on randomly assigned treatments are unbiased due to assumed exogeneity of experimental manipulations.

VARIABLES	(Experiment 1: High-Risk) Login Rate	(Experiment 2: Low-Risk) Login Rate
Active Choice	-0.115***	-0.140***
	(0.0210)	(0.0522)
Protective Default	-0.409***	-0.435***
	(0.0254)	(0.0513)
Constant	0.923***	0.797***
	(0.0116)	(0.0332)
Observations	1,526	440

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Effects of Protective Choice Architectures for Experiments 1 and 2

Table 1 provides the estimated effects for our model. Results from Experiment 1 (higher-risk setting) show that enhanced privacy controls are highly effective at decreasing consent rates. The dark default, active choice, and protective default conditions elicit login rates of 92.3%, 80.8%, and 51.4% respectively. This implies a similar 40.9% drop in login rates elicited by the protective defaults and an 11.5% drop in the active choice structure. These results provide evidence in favor of **H1**, that protective defaults are effective at combatting dark patterns.

Results from Experiment 2 (lower-risk setting) show that these effects hold, even when there is little-to-no risk to one's privacy. Those in the dark default, active choice, and protective default conditions logged in 79.7%, 65.7%, and 36.2% respectively. While the active choice structure decreases login rates by a significant 14%, there is a considerable drop elicited by the protective default amounting to a 43.5% decrease in login rates. Given that participants in Experiment 2 experience little, if any, true risk to their privacy, the estimates we find provide

compelling evidence in favor of both **H1** and **H3**, that the effects of enhanced privacy controls will persist even in low-risk settings.

3.3.3 Effect of Reversibility

We estimate the effects of reversibility using the same dependent variable used in the previous section. We find that surprisingly there are no significant baseline effects for our reversibility treatments in either experiment. The model used to estimate our results can be seen below.

$$Login_i = \beta_1 Reversible_i + \beta_2 Irreversible_i + u_i$$

$Reversible_i$ is a binary indicator for whether or not a participant was in a reversible condition. $Irreversible_i$ is a binary indicator for whether or not a participant was in an irreversible condition. We estimate this model for both experiments.

VARIABLES	(Experiment 1: High-Risk) Login Rate	(Experiment 2: Low-Risk) Login Rate
Reversibility	-0.0186	-0.00835
	(0.0269)	(0.0573)
Irreversibility	-0.0273	-0.0260
	(0.0270)	(0.0574)
Constant	0.769***	0.614***
	(0.0189)	(0.0406)
Observations	1,526	440

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Effects of Reversibility for Experiments 1 and 2

Table 2 provides the results of our estimated model. As shown, neither reversibility nor irreversibility had any significant effect. This was an initially surprising result that may have implied that the prevalence of irreversible roach motels or the subsequent regulatory

requirements for reversibility played little role in influencing decision-maker behavior, providing evidence against **H2**. However, given the strong influence of default structures on behavior shown in the previous section, we chose to estimate our model with the goal of uncovering interaction effects between reversibility and defaults.

Simply comparing the login rates across conditions provides interesting initial results. For example, in Experiment 1 (higher-risk), we see the surprising trend that informing the participant that the choice is either reversible *or* irreversible enhances the effect of the protective default structure, further lowering login rates. We find that adding reversibility or irreversibility decreases login rates from 58.3% to 47.8% and 47.9% respectively. In Experiment 2 (lower-risk) we see that reversibility actually enhances the effects of *either* default structure. In other words, in the dark default treatment where login rates are roughly 25% higher than in the protective default conditions absent any reversibility information, adding reversibility or irreversibility increases login rates from 70% to 89.1% and 80.8% respectively. Likewise, in the protective default condition, reversibility and irreversibility decrease login rates from 44.9% to a shockingly low 28.8% and 35.3% respectively.

Table 3 provides the estimates of our reversibility treatment effects across choice architectures, confirming our initial findings above. Columns 1 through 3 provide estimated effects of reversibility and irreversibility in the higher-risk setting for those in a dark default condition (DD), an active choice condition (AC), and a protective default condition (PD), respectively. Columns 4 through 6 provide similar estimates for the lower-risk setting.

In the higher-risk settings, reversibility had a significant effect for those in the protective default condition, lowering login rates by 10.5%. Given the already high-stakes of the setting in Experiment 1, it may be that we elicited a similar phenomenon to Peer and Acquisti (2016)) in

which informing the decision-maker about the reversibility of the decision increased the perceived-stakes of the decision. In a lower-risk setting, we find that reversibility enhances the effects of defaults in either direction with a significant 19% increase and 16% decrease in login rates for the dark default and protective default conditions respectively. Interestingly, we find no statistically significant effects of irreversibility across default structures.

	Experiment 1: High-Risk			Experiment 2: Low-Risk		
	(1: DD)	(2: AC)	(3: PD)	(4: DD)	(5: AC)	(6: PD)
VARIABLES	Login Rate	Login Rate	Login Rate	Login Rate	Login Rate	Login Rate
Reversibility	0.000878	0.0179	-0.105*	0.191**	-0.0222	-0.161*
	(0.0286)	(0.0433)	(0.0551)	(0.0802)	(0.0964)	(0.0958)
Irreversibility	3.29e-05	-0.00247	-0.105*	0.108	-0.0957	-0.0960
	(00287)	(0.0436)	(0.0547)	(0.0856)	(0.101)	(0.0986)
Constant	0.923***	0.802***	0.583***	0.700***	0.696***	0.449***
	(0.0207)	(0.0314)	(0.0382)	(0.0655)	(0.0686)	(0.0718)
Observations	532	504	490	148	140	152

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Effects of Reversibility for Experiments 1 and 2 Across Choice Architectures

Given that privacy regulation generally encourages or requires a combination of protective defaults and reversibility, this is concerning for firms engaging in lower-risk data collection. It implies that when providing the encouraged enhanced privacy protections to their customers, rates of consent may drop dramatically given this interaction effect (as shown by the 28.8% login rate described above). We find partial support for **H2** where reversibility lowered rates of consent when paired with a protective choice default. Again, these results provide further evidence in favor of **H3**, showing that enhanced privacy controls significantly effect consumer behavior even when the risk to privacy is lowered.

Table 4 below estimates the full model for Experiments 1 and 2, including interaction terms between each of our distinct factors. We find in Experiment 1 that a protective default

alone decreases login rates by 33.9% but by adding reversibility login rates decrease by an additional 10.5%. Similarly, the interaction between a protective default and reversibility is exceedingly strong in the lower-risk setting. In this setting, the protective default alone decreases login rates by 25.1% but adding reversible consent decreases the login rate by an *additional* 16.1%. This provides evidence in favor of **H3** that the effects of enhanced privacy protections persist without regard to the risk level of the decision context.

	Experiment 1: High-Risk	Experiment 2: Low-Risk
	(1)	(2)
VARIABLES	Login Rate	Login Rate
Active Choice	-0.120***	-0.00435
	(0.0376)	(0.0948)
Protective Default	-0.339***	-0.251**
	(0.0434)	(0.0972)
Reversibility	0.000878	0.191**
	(0.0286)	(0.0802)
Irreversibility	3.29e-05	0.108
	(0.0287)	(0.0857)
AC × Reversibility	0.0170	-0.213*
	(0.0518)	(0.125)
AC × Irreversibility	-0.00250	-0.203
	(0.0522)	(0.132)
PD × Reversibility	-0.106*	-0.352***
	(0.0620)	(0.125)
PD × Irreversibility	-0.105*	-0.204
	(0.0618)	(0.131)
Constant	0.923***	0.700***
	(0.0207)	(0.0655)
Observations	1,526	440

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Estimates with Interaction Terms for Experiments 1 and 2

4. Experiment 3

Experiment 3 differs from our first two experiments in several ways. First, we increase the realism of the decision context by using the context of third-party data sharing through

mobile applications. Participants believe they are completing a screening survey to take part in a mobile application pilot program. Participants are told that, if selected, they would be required to download a mobile application to their personal device and use it for a period of three weeks. At the end of the survey, we ask them to consent to our sharing of their data with third party corporate sponsors to aide in the development of future apps. Additionally, while Experiments 1 and 2 were conducted at separate times with different samples, Experiment 3 uses one sample and varies the stakes of decision context as an exogenous manipulation. This allows us to directly compare the effects of enhanced privacy protections across levels of risk.

4.1 Experimental Design

We utilize a four factor between-subjects online experiment for Experiment 3. Our manipulated factors differentiate 1) the choice structure of the consent decision (dark default, active choice, or protective default), 2) the reversibility of the consent (explicitly reversible or no reversibility statement[12]), 3) the presence of a social nudge intended to temper the effects of enhanced privacy protections in the low-risk setting and have little to no effect in the high risk setting and 4) the privacy risk of the decision (lower- or higher-risk). Our first two factors are structured similarly to the first two experiments. Similarly, in the reversibility conditions, participants are informed that they can change their decision at any time by contacting the research team. Consent structures can be found in Appendix F.

4.1.2 Social Norm Nudge

Participants who were assigned to a nudge treatment were informed that 60% of past participants chose to share their data with third parties (see Appendix F). We wanted the metric

[12] We chose to not include the explicitly irreversible condition that we used in Experiments 1 and 2 given the lack of a significant difference in effect between reversibility and irreversibility.

used to be believable and so we arrived at the number by asking participants in a pilot study to estimate what percent of people they *believed* would consent to the choice. The average response was roughly 60%.

4.1.3 Risk Level

We manipulate the level of risk across conditions by varying the requested permissions of the mobile application. Participants are told that the application they will be testing is a news and entertainment app named "Newsvia." They are shown screenshots of the app's home page and account creation process (see Appendix G) before being directed to images of the app requesting certain permissions from the user. We determined the perceived risk of various combinations of app permissions using a variation of the measurement tool developed by Gu et al. (2017)). The full list of questions can be found in the Appendix H. We chose permissions that were perceived to be either high- or low-risk while controlling for the perceived usefulness of the app. The average response to each question across the two conditions is shown in Appendix H.

The application presented in the lower-risk conditions asked for the following permissions: the ability to send notifications, the ability to connect to Wi-Fi networks, and the ability to create a shortcut to the app on the user's home screen. For the higher-risk conditions, the permissions used included access to the device's location, the ability to access the device's microphone and record audio, and the ability to send and view SMS messages. Images shown to the participants can be found in Appendix G.

4.1.3 Participants and Procedures

2200 participants were recruited from Amazon Mechanical Turk to take part in this experiment. After excluding participants that gave unengaged responses or asked for their data to

be removed after they were debriefed, we had a final sample size of 2069. Participants were compensated $0.75 for completing the study. They began by reading a brief introduction to the application. Then they completed a demographic profile and answered questions relating to their mobile device, to increase the believability of the task.

After completing the preliminary questions, they were shown the application renderings described above. Then they were directed to the primary outcome in our study which asked them to consent to the researchers sharing their data from the application testing with third-party corporate sponsors. The choice was structured based on their default, reversibility, and social nudge conditions. After making their decision, we asked them which of the three permissions they were planning on granting the application during testing. They were informed that their answers would not impact their chances of being selected for the pilot. Finally, participants answered a series of exit questions and were debriefed.

4.2 Results

As in the first two experiments, we evaluate observable variables for each participant (age, race, gender, education, and employment) and find balance across most variables in both experiments. Descriptive statistics and a table of pairwise comparisons can also be found in the Appendix E. The results of our analysis show that 1) default effects may actually be stronger in low-risk settings, 2) introducing reversibility has little impact on consent rates in high-risk settings but significantly lowers consent rates in low-risk settings, particularly when paired with a protective default, and 3) social norm nudges have the surprising effect of *lowering* consent rates across conditions implying an upstream backlash effect that further shows the difficulty of tempering the unintended effects of regulatory privacy changes.

4.3.2 Effect of Defaults

Table 5 provides the mean consent rate for each of the default treatments in both the lower- and higher-risk settings. Interestingly, we find that the dark default has a *stronger* effect in the lower-risk setting where consent rates were almost 84% (a 13.8% increase from the active choice condition) compared to 76% (a 9% increase from the active choice condition) in the higher-risk setting. This implies that consumers are able to recognize the risk of the decision setting and are less impacted by dark defaults when the risk is high. This is an important result showing that absent enhanced privacy controls, consumers are able to appropriately respond to the use of some dark patterns without intervention.

Further, we find that the protective default has a strong effect on consent rates in both the lower- and higher-risk settings decreasing consent rates by 42.3% and 38.6% respectively compared to the active choice condition. Interestingly, there is little difference in the consent rate between the lower- and higher-risk settings for those given an active choice structure. This implies that participants did not significantly change their consent preferences even in light of the risk to their privacy being lowered. These results again provide evidence in favor of **H1** and **H3**.

TREATMENT	Dark Default	Active Choice	Protective Default
Lower-Risk	83.9%	70.1%	27.8%
Higher-Risk	76.0%	67.0%	28.4%

Table 5: Experiment 3 Consent Rates

The dependent variable in our regression is a binary indicator variable that represents whether or not the participant *i* chose to consent to sharing their data with third parties.

$$Consent_i = \beta_1 ActiveChoice_i + \beta_2 ProtectiveDefault_i + \beta_3 LowerRisk_i + \beta_4 (ActiveChoice_i \times LowerRisk_i) + \beta_5 (ProtectiveDefault_i \times LowerRisk_i) + u_i$$

$ActiveChoice_i$ and $ProtectiveDefault_i$ are the same binary indicators used in earlier analyses. $LowerRisk_i$ is a binary indicator for whether or not a participant was in the lower-risk condition. We also include interaction terms between the risk level and the consent structure to measure differences between the higher- and lower-risk settings.

As a reminder, in the first two experiments, we found that default effects persisted across risk-levels. This was a concerning result given that even in a decision context with virtually no privacy risk, participants were still influenced by changes required most privacy regulation. Table 6 below provides default effects for the higher- and lower-risk conditions separately as well as the combined model specified above. As the structure of Experiment 3 allows us to compare effects across risk levels, these estimates imply an even more concerning result that the effects of protective choice architectures are actually *higher* in low-risk settings. Those given a protective default structure decreased consent rates by 47.6% and 56.1% in high- and low-risk settings respectively, an 8.5% increase in the effectiveness of the protective default. While **H3** simply anticipates that the effects of privacy controls will persist across risk levels, these findings go further to show that they may actually be stronger in low-risk settings. The results also confirm that in the dark default conditions consent rates were higher in the lower-risk setting, showing that participants appropriately responded to changes in risk when presented with a dark default.

VARIABLES	(1: High-Risk) Consent	(2: Low-Risk) Consent	(3: Combined) Consent
Active Choice (AC)	-0.0905***	-0.138***	-0.0905***
	(0.0344)	(0.0316)	(0.0344)
Protective Default (PD)	-0.476***	-0.561***	-0.476***
	(0.0334)	(0.0314)	(0.0334)
Low-Risk			0.0788**
			(0.0307)

AC × Low-Risk			-0.0473
			(0.0467)
PD × Low-Risk			-0.0853*
			(0.0459)
Constant	0.760***	0.839***	0.760***
	(0.0230)	(0.0203)	(0.0230)
Observations	1,033	1,036	2,069

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Effects of Protective Choice Architectures for Experiment 3

4.3.3 Effect of Reversibility

We again use our above model to estimate the effects of reversibility for Experiment 3. Given the findings from the first two experiments, we estimate effects across choice architectures. Table 7 reports the results. Interestingly, we see little impact of reversibility on consent choices in the higher-risk setting. We do confirm our findings from the first two experiments that reversibility further increases consent rates when paired with a dark default in a lower-risk setting. Overall, we find mixed findings in regards to the impact of reversibility across our three experiments. In the context of Experiment 3, it appears as if the default structure, not reversibility, drives the majority of behavioral changes across risk levels.

VARIABLES	(1: DD) Consent	(2: AC) Consent	(3: PD) Consent
Reversibility	0.0169	0.0178	-0.0287
	(0.0461)	(0.0512)	(0.0486)
Lower-Risk	0.0849*	0.0834*	-0.0192
	(0.0441)	(0.0493)	(0.0488)
Reversibility × Lower-Risk	-0.0118	-0.101	0.0239
	(0.0614)	(0.0703)	(0.0684)
Constant	0.751***	0.661***	0.299***
	(0.0333)	(0.0363)	(0.0355)
Observations	675	697	697

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Reversibility Effects for Experiment 3

4.3.4 Effect of Social Norm Nudge

We estimate the effects of our social norm nudge on consent rates for both higher- and lower-risk decision contexts. Table 8 below provides our results.

VARIABLES	(1: Low-Risk) Consent	(2: High-Risk) Consent	(3: Combined) Consent
Social Norm Nudge (SNN)	-0.0419	-0.0572*	-0.0572*
	(0.0304)	(0.0308)	(0.0308)
Low-Risk			0.0248
			(0.0307)
SNN × Low-Risk			0.0153
			(0.0433)
Constant	0.624***	0.599***	0.599***
	(0.0217)	(0.0217)	(0.0217)
Observations	1,036	1,033	2,069

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 8: Social Norm Nudge Effects

The original intention for the nudge was to increase consent rates in a lower-risk setting while having no effect in a higher-risk setting. Interestingly, we find that the nudge decreases consent rates in both the lower- and higher-risk settings, although only to a statistically significant amount in the latter. This could imply that participants viewed the nudge as manipulative regardless of risk, and it elicited a backlash effect that led to lower rates of consent.

VARIABLES	(1: DD) Consent	(2: AC) Consent	(3: PD) Consent
Social Norm Nudge	-0.0927**	-0.0806	0.0124
	(0.0457)	(0.0510)	(0.0485)
Lower-Risk	0.0563	-0.00720	0.0268
	(0.0407)	(0.0492)	(0.0499)
Nudge × Lower-Risk	0.0449	0.0772	-0.0632
	(0.0611)	(0.0704)	(0.0684)

Constant	0.807***	0.710***	0.278***
	(0.0303)	(0.0350)	(0.0346)
Observations	675	697	697

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 9: Social Norm Nudge Effects across Choice Architectures

Table 9 provides the estimated effect of the nudge across different choice architectures. We find that in both the dark default and active choice conditions, the nudge decreased consent rates by 9.3% and 8.1% respectively, although slightly insignificant for the active choice condition (p=0.115). Our interaction term shows no significant difference in these effects between the higher- and lower-risk settings. Interestingly, we see no effect of our nudge the protective default conditions. This provides evidence against **H4** and highlights the difficulty in counteracting the social costs of privacy regulation in settings with little privacy risk.

5. Discussion & Conclusion

Our findings provide significant insight into the unintended consequences of enhanced privacy protections. First, we provide substantial evidence that manipulating choice architecture through default structures has a massive effect on consent decisions. Protective defaults lower consent rates by up to 56% in our setting. Importantly, we show that these effects persist even when privacy faces a considerably lower risk. Additionally, we show the ability of pairing an active choice structure or protective defaults with reversibility to amplify the default effect. For example, in Experiment 1, participants in the protective default condition with no added reversibility treatments logged in 45% of the time. Conversely, those who saw the protective default and reversibility logged in only 29% of the time.

Importantly, in Experiment 3 we show that the effects of these regulatory changes may actually be *stronger* in lower-risk settings. For example, while an active choice structure and a

protective default decrease consent rates by 9% and 48% respectively in the high-risk setting, they decrease consent by 14% and 56% respectively in the low-risk setting. One possible explanation for this is that cognitive biases drive much of the choice architecture effects. In higher-risk settings, individuals may be more on guard and less susceptible to those biases, leading to lowered default effects. These findings again highlight the mechanisms through which a policy overreaction may harm both firms and consumers.

Finally, we show the difficulty in tempering these unintended effects. We introduced a social norm nudge intended to alleviate privacy concerns in the lower-risk setting leading to increased rates of consent without eliciting any change in a higher-risk setting. We instead elicited a backlash effect that further decreased consent rates, specifically in a higher-risk setting. This implies that consumers may be sensitive to the use of certain nudges (including dark patterns) and react by changing subsequent privacy choices.

This work is not without limitations. First, while we increase the external validity of our findings by conducting three experiments across two different privacy contexts, in reality, our participants' privacy was not truly at risk in any of the experiments. While we did not conduct a field experiment or collect secondary data, the structure of our third experiment specifically allowed for participants to *believe* that their privacy was truly at risk and therefore their responses should be indicative of how they would act in the real world. Further, while most work that considers the impact of GDPR or other policy changes focuses on the sweeping effects of the entire change, we focus only on changes required of consent elicitation when using it as the legal bases for data processing. While this may limit the reach of our findings, it also allows for a more thorough understanding of the contexts that we describe.

Despite these limitations, our work opens numerous doors for future research. Concerning the positive and negative effects of modern privacy regulation, we show that the effects of enhanced privacy controls persist, or are even stronger, when an individual's privacy faces little risk. This is an essential first step to addressing potential externalities that arise from privacy regulation. Further, we show that these unintended costs are often difficult to overcome. By utilizing nudging to help consumers contextualize risk, we instead elicit a backlash effect that drove consent further down. This opens up a potential avenue for research that explores other nudges or interventions that may be successful in addressing the issue of risk in consumer privacy settings.

Our work highlights the importance of enhanced privacy controls in high-risk settings while showing the mechanisms through which these regulations may be an overreaction in lower-risk settings. By introducing the framework of a policy overreaction to this literature, we show both the considerable importance of privacy regulation and the consequences of unchecked dark patterns while also acknowledging the potential unintended consequences that may accompany such regulation. By considering heterogeneous effects of privacy regulation around levels of risk, we show that these unintended consequences are non-trivial and require further consideration. While level of privacy risk is one source of heterogeneity, other sources may provide fruitful future research endeavors including societal importance of the data collection efforts or user familiarity with the privacy context.

Additionally, regulation like GDPR is often broad, covering a sweeping array of privacy-related issues. Isolating specific tenets of said regulation will further lend to the insights surrounding its enactment. However, these tenets do not exist in a vacuum and likely interact, adding further complexity to their effects. Addressing the effects of privacy-regulation from both

the perspective of individual changes and the interaction of such changes will allow for further development of the both the academic literature and policy discussions. Possible examples could include the interaction between data security requirements and privacy by design or the effects of net neutrality (which classifies internet service providers as telecommunications services instead of information services, thus dividing privacy regulation enforcement between the FCC and the FTC[13]) on the right to be forgotten.

Our work also has implications for the choice architecture literature; specifically work that addresses the role of choice architecture in privacy decision making. Prior work has laid the groundwork for analyzing individual manipulations to choice architecture such as the ability to compare privacy risks side-by-side when choosing products (Egelman et al. 2013), effectively opting-out of a data sharing decision (Cho et al. 2019), and reversible consent decisions (Peer and Acquisti 2016). Given the interaction between varying choice architectures utilized by firms, we explore their effects in conjunction. For example, our findings show that despite reversibility having little effect on behavior at a baseline, by interacting its use with varying choice defaults, we can better understand how this mechanism influences decision makers. Given that the directional effect of reversibility aligns with the effect of the default it is paired with, it is implied that reversibility, despite not having an effect on its own, enhances the effects of choice defaults. By exploring the interaction of varying choice architecture tools, future research can uncover additional relationships and nuances to the literature.

Prior work has often highlighted the role of choice architecture in manipulating consumers into making poorer privacy decisions and, importantly, introduced more consumer-oriented

[13] https://www.govtech.com/policy/gao-report-the-net-neutrality-debate-complicates-data-privacy.html

choice architectures that can counter those effects. These could include choosing more lenient disclosure settings (Acquisti et al. 2012; Adjerid et al. 2019), reducing privacy risk perceptions (Tsay-Vogel et al. 2018; Xu et al. 2009), or, in even more recent work, the inability to maintain personal boundaries in online social interactions (Zhang et al. 2022). Our work extends these findings by examining any unintended consequences these protective choice architectures may have. Our results show that the role of choice architecture in policy overreactions is non-trivial and by addressing it (e.g., incorporating heterogeneity around regulatory requirements for differing levels of privacy risk), the social costs of these policies may be reduced, opening the door for numerous explorations of the unintended consequences of consumer-oriented choice architecture changes.

Our results also contribute to the nascent literature concerning dark patterns and online privacy. Notably, consumer awareness of dark patterns (Maier and Harr 2020) and their overall privacy concerns[14] have grown considerably in the decade since the first warning signal was shone on the use of dark patterns. Therefore, consumer interactions with dark patterns and their effectiveness in manipulating user choice are likely to change as these trends continue. Our findings provide evidence showing that the risk of consumer backlash to dark patterns is salient. Specifically, we show that irreversibility and social norm nudges (the latter of which was intended to be a positive nudge in our work but is used extensively as a dark pattern in the wild) may both elicit backlash from consumers that lead to more privacy protective decisions upstream. Future work should continue to explore the role of dark patterns in quickly changing online domains such as social media, crowdsourced work, or online health communities.

[14] https://www.techrepublic.com/article/data-privacy-is-a-growing-concern-for-more-consumers/

This is not to say that the effectiveness of dark patterns is not a concern. We provide clear evidence that some dark patterns (most notably, dark defaults) are still highly effective at influencing consumers. However, no work, to our knowledge, considering dark patterns has looked at their effectiveness across levels of risk. We find, surprisingly, that dark patterns may have a stronger effect in lower-risk settings. This could imply that in higher risk settings, consumers may be less impacted by the cognitive biases that lead to, for instance, default effects and therefore make choices that are more aligned with their true preferences (moving towards the active choice level of consent). Further work is needed to definitively conclude why we find this heterogeneity of dark pattern effects across risk levels.

Finally, this work introduces privacy as a potential domain for policy overreaction in the political science literature. Given the often large and all-encompassing nature of privacy regulation, it is likely that this is an apt context to explore. GDPR, specifically, fits largely into the "universal policy" classification provided as one of the key sources of policy overreaction (Maor 2019). Additionally, given the prevalence of trigger events that lead to heightened demand for privacy regulation, the risk of an overreaction is non-trivial (Jennings et al. 2020). By utilizing this theoretical perspective in our work, we are able to provide concrete examples of where policy overreactions could stem from within the privacy setting and open the door for further cross-discipline explorations of the phenomenon.

We also provide significant implications for policymakers and industry leaders. While industry professionals and other concerned parties have cited evidence showing the downsides of GDPR and the like, we provide the first evidence that highlights the consequences of a policy overreaction resulting from heterogenous effects around levels of risk. These findings can advance efforts to adopt a more balanced approach to consumer privacy protections. Importantly,

these results are perhaps most insightful to policymakers. The balance that we highlight between privacy protection and significant unintended consequences of said protection is essential to crafting future regulation. Most significantly, we highlight the need for differing levels of privacy protection for different levels of privacy risk. In addition, our results highlight the potential for a significant consumer reaction to a privacy-driven policy overreaction that enhances privacy protections without concern for levels of privacy risk.

6. References

Acquisti, A., and Grossklags, J. 2005. "Privacy and Rationality in Individual Decision Making," *Security & Privacy, IEEE* (3), pp. 26-33.

Acquisti, A., John, L. K., and Loewenstein, G. 2012. "The Impact of Relative Standards on the Propensity to Disclose," *Journal of Marketing Research* (49:2), pp. 160-174.

Acquisti, A., Sleeper, M., Wang, Y., Wilson, S., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L., Komanduri, S., Leon, P., Sadeh, N., and Schaub, F. 2017. "Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online," *ACM Computing Surveys* (50), pp. 1-41.

Adjerid, I., Acquisti, A., and Loewenstein, G. 2019. "Choice Architecture, Framing, and Cascaded Privacy Choices," *Management Science* (65:5), pp. 2267-2290.

Adjerid, I., Acquisti, A., Telang, R., Padman, R., and Adler-Milstein, J. 2015. "The Impact of Privacy Regulation and Technology Incentives: The Case of Health Information Exchanges," *Management Science* (62:4), pp. 1042-1063.

Aridor, G., Che, Y.-K., and Salz, T. 2021. "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from Gdpr," *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 93-94.

Article 29, T. W. P. 2017. "Guidelines on Consent under Regulation 2016/679."

Baroni, L. A., Puska, A. A., de Castro Salgado, L. C., and Pereira, R. 2021. "Dark Patterns: Towards a Socio-Technical Approach," *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*, pp. 1-7.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., and Galing, S. 2017. "Should Governments Invest More in Nudging?," *Psychological Science* (28:8), pp. 1041-1055.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. 2015. "The Effect of Providing Peer Information on Retirement Savings Decisions," *The Journal of finance* (70:3), pp. 1161-1201.

Bhargava, S., and Manoli, D. 2015. "Psychological Frictions and the Incomplete Take-up of Social Benefits: Evidence from an Irs Field Experiment," *American Economic Review* (105:11), pp. 3489-3529.

Birkland, T. A. 2006. *Lessons of Disaster: Policy Change after Catastrophic Events*. Georgetown University Press.

Bleier, A., Goldfarb, A., and Tucker, C. 2020. "Consumer Privacy and the Future of Data-Based Innovation and Marketing," *International Journal of Research in Marketing* (37:3), pp. 466-480.
Blumenthal, M., Christian, C., and Slemrod, J. 2001. "Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota," *National Tax Journal* (54:1), pp. 125-138.
Bösch, C., Erb, B., Kargl, F., Kopp, H., and Pfattheicher, S. 2016. "Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns," *Proceedings on Privacy Enhancing Technologies* (2016:4), pp. 237-254.
Brignull, H. 2022. "Deceptive Designs." from deceptive.design
Brown, C. L., and Krishna, A. 2004. "The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice," *Journal of Consumer Research* (31:3), pp. 529-539.
Butterworth, M. 2018. "The Ico and Artificial Intelligence: The Role of Fairness in the Gdpr Framework," *Computer Law & Security Review* (34:2), pp. 257-268.
Camerer, C., Issacharoff, S., Loewenstein, G., o Donoghue, T., and Rabin, M. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'," *Law & Economics*).
Cara, C. 2019. "Dark Patterns in the Media: A Systematic Review," *Network Intelligence Studies* (7:14), pp. 105-113.
Cho, H., Roh, S., and Park, B. 2019. "Of Promoting Networking and Protecting Privacy: Effects of Defaults and Regulatory Focus on Social Media Users' Preference Settings," *Computers in Human Behavior* (101), pp. 1-13.
Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. 2003. "Optimal Defaults," *American Economic Review* (93:2), pp. 180-185.
Cialdini, R. B., Kallgren, C. A., and Reno, R. R. 1991. "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior," in *Advances in Experimental Social Psychology*. Elsevier, pp. 201-234.
Custers, B., Dechesne, F., Sears, A. M., Tani, T., and van der Hof, S. 2018. "A Comparison of Data Protection Legislation and Policies across the Eu," *Computer Law & Security Review* (34:2), pp. 234-243.
Dinev, T., and Hart, P. 2006. "An Extended Privacy Calculus Model for E-Commerce Transactions," *Information Systems Research* (17:1), pp. 61-80.
Dinner, I., Johnson, E. J., Goldstein, D. G., and Liu, K. 2011. "Partitioning Default Effects: Why People Choose Not to Choose," *J Exp Psychol Appl* (17:4), pp. 332-341.
Egelman, S., Felt, A., and Wagner, D. 2013. "Choice Architecture and Smartphone Privacy: There's a Price for That,").
Erceg-Hurn, D. M., and Steed, L. G. 2011. "Does Exposure to Cigarette Health Warnings Elicit Psychological Reactance in Smokers?," *Journal of applied social psychology*).
Evans, A., Dillon, K., Goldin, G., and Krueger, J. 2011. "Trust and Self-Control: The Moderating Role of the Default," *Judgment and Decision Making* (6), pp. 697-705.
Fiske, S. T., and Taylor, S. E. 1991. *Social Cognition, 2nd Ed*. New York, NY, England: Mcgraw-Hill Book Company.
Frey, B. S., and Meier, S. 2004. "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment," *American Economic Review* (94:5), pp. 1717-1722.

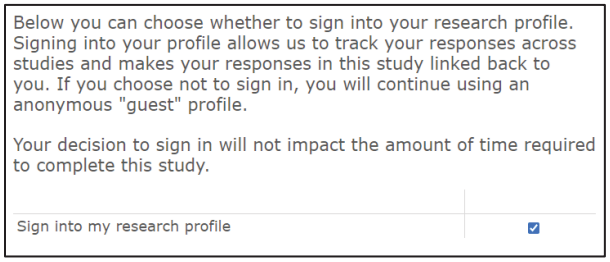
FTC. 2020. "Regarding Dark Patterns in the Matter of Age of Learning, Inc.," in: *1723186,* F.T. Comission (ed.).
Gal, M. S., and Aviv, O. 2020. "The Competitive Effects of the Gdpr," *Journal of Competition Law & Economics* (16:3), pp. 349-391.
Gerber, A. S., and Rogers, T. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You," *The Journal of Politics* (71:1), pp. 178-191.
Godinho de Matos, M., and Adjerid, I. 2021. "Consumer Consent and Firm Targeting after Gdpr: The Case of a Large Telecom Provider," *Management Science*).
Goecks, J., Edwards, W. K., and Mynatt, E. D. 2009. "Challenges in Supporting End-User Privacy and Security Management with Social Navigation," *Proceedings of the 5th Symposium on Usable Privacy and Security*, pp. 1-12.
Goecks, J., and Mynatt, E. D. 2005. "Supporting Privacy Management Via Community Experience and Expertise," in *Communities and Technologies 2005*. Springer, pp. 397-417.
Goes, P. B. 2013. "Editor's Comments: Information Systems Research and Behavioral Economics," *MIS Q.* (37:3), pp. iii–viii.
Goldberg, S., Johnson, G., and Shriver, S. K. 2019. "Regulating Privacy Online: The Early Impact of the Gdpr on European Web Traffic & E-Commerce Outcomes."
Goldfarb, A., and Tucker, C. 2011. "Privacy Regulation and Online Advertising," *Management Science* (57:1), pp. 57-71.
Goswami, I., and Urminsky, O. 2016. "When Should the Ask Be a Nudge? The Effect of Default Amounts on Charitable Donations," *Journal of Marketing Research* (53:5), pp. 829-846.
Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., and Toombs, A. L. 2018. "The Dark (Patterns) Side of Ux Design," *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1-14.
Greenberg, S., Boring, S., Vermeulen, J., and Dostal, J. 2014. "Dark Patterns in Proxemic Interactions: A Critical Perspective," *Proceedings of the 2014 conference on Designing interactive systems*, pp. 523-532.
Gu, J., Xu, Y. C., Xu, H., Zhang, C., and Ling, H. 2017. "Privacy Concerns for Mobile App Download: An Elaboration Likelihood Model Perspective," *Decision Support Systems* (94), pp. 19-28.
Huh, Y. E., Vosgerau, J., and Morewedge, C. K. 2014. "Social Defaults: Observed Choices Become Choice Defaults," *Journal of Consumer Research* (41:3), pp. 746-760.
Janssen, R., Kesler, R., Kummer, M. E., and Waldfogel, J. 2022. "Gdpr and the Lost Generation of Innovative Apps," National Bureau of Economic Research.
Jennings, W., Farrall, S., Gray, E., and Hay, C. 2020. "Moral Panics and Punctuated Equilibrium in Public Policy: An Analysis of the Criminal Justice Policy Agenda in Britain," *Policy Studies Journal* (48:1), pp. 207-234.
Jia, J., Jin, G. Z., and Wagman, L. 2021. "The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment," *Marketing Science* (40:4), pp. 661-684.
John, P. C. H. 2018. "How Best to Nudge Taxpayers?: The Impact of Message Simplification and Descriptive Social Norms on Payment Rates in a Central London Local Authority," *Journal of Behavioral Public Administration* (1:1), pp. 1-11.
Johnson, E., Shu, S., Dellaert, B., Fox, C., Goldstein, D., Häubl, G., Larrick, R., Payne, J., Peters, E., Schkade, D., Wansink, B., and Weber, E. 2012. "Beyond Nudges: Tools of a Choice Architecture," *Marketing Letters* (23:2), pp. 487-504.

Johnson, E. J., Bellman, S., and Lohse, G. L. 2002. "Defaults, Framing and Privacy: Why Opting in-Opting Out1," *Marketing Letters* (13:1), pp. 5-15.
Johnson, E. J., and Goldstein, D. 2003. "Do Defaults Save Lives?," *Science* (302:5649), pp. 1338-1339.
Johnson, G., Shriver, S., and Goldberg, S. 2022. "Privacy & Market Concentration: Intended & Unintended Consequences of the Gdpr," *Available at SSRN 3477686*).
Keller, P. A., Harlam, B., Loewenstein, G., and Volpp, K. G. 2011. "Enhanced Active Choice: A New Method to Motivate Behavior Change," *Journal of Consumer Psychology* (21:4), pp. 376-383.
Lauer, T. W., and Deng, X. 2007. "Building Online Trust through Privacy Practices," *International Journal of Information Security* (6:5), pp. 323-331.
Layton, R. 2019. *10 Problems of the Gdpr: The Us Can Learn from the Eu's Mistakes and Leapfrog Its Policy*. American Enterprise Institute.
Levav, J., Heitmann, M., Herrmann, A., Iyengar, S., xa, and S. 2010. "Order in Product Customization Decisions: Evidence from Field Experiments," *Journal of Political Economy* (118:2), pp. 274-299.
Li, H., Sarathy, R., and Zhang, J. 2008. "The Role of Emotions in Shaping Consumers' Privacy Beliefs About Unfamiliar Online Vendors," *Journal of Information privacy and Security* (4:3), pp. 36-62.
Luguri, J., and Strahilevitz, L. J. 2021. "Shining a Light on Dark Patterns," *Journal of Legal Analysis* (13:1), pp. 43-109.
Maier, M., and Harr, R. 2020. "Dark Design Patterns: An End-User Perspective," *Human Technology* (16:2), p. 170.
Maor, M. 2012. "Policy Overreaction," *Journal of Public Policy* (32:3), pp. 231-259.
Maor, M. 2019. "Strategic Policy Overreaction as a Risky Policy Investment," *International Review of Public Policy* (1:1: 1), pp. 46-64.
Maor, M. 2021. "Deliberate Disproportionate Policy Response: Towards a Conceptual Turn," *Journal of Public Policy* (41:1), pp. 185-208.
Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., and Narayanan, A. 2019. "Dark Patterns at Scale: Findings from a Crawl of 11k Shopping Websites," *Proceedings of the ACM on Human-Computer Interaction* (3:CSCW), pp. 1-32.
McKenzie, C. R., Liersch, M. J., and Finkelstein, S. R. 2006. "Recommendations Implicit in Policy Defaults," *Psychol Sci* (17:5), pp. 414-420.
Miller, A. R., and Tucker, C. 2017. "Privacy Protection, Personalized Medicine, and Genetic Testing," *Management Science* (64:10), pp. 4648-4668.
Mol, J. M., Botzen, W. W., Blasch, J. E., Kranzler, E. C., and Kunreuther, H. C. 2021. "All by Myself? Testing Descriptive Social Norm-Nudges to Increase Flood Preparedness among Homeowners," *Behavioural Public Policy*), pp. 1-33.
Mullainathan, S., and Thaler, R. H. 2000. "Behavioral Economics," *National Bureau of Economic Research Working Paper Series* (No. 7948).
Nahmias, Y., Perez, O., Shlomo, Y., and Stemmer, U. 2019. "Privacy Preserving Social Norm Nudges," *Mich. Tech. L. Rev.* (26), p. 43.
Park, C., Jun, S., and Macinnis, D. 2000. "Choosing What I Want Versus Rejecting What I Do Not Want: An Application of Decision Framing to Product Option Choice Decisions," *Journal of Marketing Research - J MARKET RES-CHICAGO* (37), pp. 187-202.

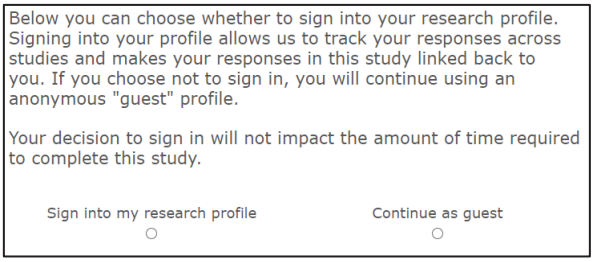
Patil, S., Page, X., and Kobsa, A. 2011. "With a Little Help from My Friends: Can Social Navigation Inform Interpersonal Privacy Preferences?," *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 391-394.
Peer, E., and Acquisti, A. 2016. "The Impact of Reversibility on the Decision to Disclose Personal Information," *Journal of Consumer Marketing* (33), pp. 428-436.
Quigley, M. 2013. "Nudging for Health: On Public Policy and Designing Choice Architecture," *Medical law review* (21:4), pp. 588-621.
Rottenstreich, Y., and Hsee, C. K. 2001. "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk," *Psychological science* (12:3), pp. 185-190.
Ruohonen, J., and Hjerppe, K. 2022. "The Gdpr Enforcement Fines at Glance," *Information Systems* (106), p. 101876.
Samuelson, W., and Zeckhauser, R. 1988. "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty* (1:1), pp. 7-59.
Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science* (18:5), pp. 429-434.
Sunstein, C. R., and Zeckhauser, R. 2011. "Overreaction to Fearsome Risks," *Environmental and Resource Economics* (48:3), pp. 435-449.
Thaler, R. H., and Sunstein, C. R. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT, US: Yale University Press.
Thaler, R. H., Sunstein, C. R., and Balz, J. P. 2013. "Choice Architecture," in *The Behavioral Foundations of Public Policy.* Princeton, NJ, US: Princeton University Press, pp. 428-439.
Tsai, J. Y., Egelman, S., Cranor, L., and Acquisti, A. 2011. "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research* (22:2), pp. 254-268.
Tsay-Vogel, M., Shanahan, J., and Signorielli, N. 2018. "Social Media Cultivating Perceptions of Privacy: A 5-Year Analysis of Privacy Attitudes and Self-Disclosure Behaviors among Facebook Users," *New media & society* (20:1), pp. 141-161.
Tversky, A., and Kahneman, D. 1974. "Judgment under Uncertainty: Heuristics and Biases," *Science* (185:4157), pp. 1124-1131.
Waldman, A. E. 2020. "Cognitive Biases, Dark Patterns, and the 'Privacy Paradox'," *Current opinion in psychology* (31), pp. 105-109.
Whitley, E. A. 2009. "Informational Privacy, Consent and the "Control" of Personal Data," *Inf. Secur. Tech. Rep.* (14:3), pp. 154–159.
Wood, S. L. 2001. "Remote Purchase Environments: The Influence of Return Policy Leniency on Two-Stage Decision Processes," *Journal of Marketing Research* (38:2), pp. 157-169.
Xu, H., Teo, H.-H., Tan, B. C., and Agarwal, R. 2009. "The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services," *Journal of management information systems* (26:3), pp. 135-174.
Zaeem, R. N., and Barber, K. S. 2020. "The Effect of the Gdpr on Privacy Policies: Recent Progress and Future Promise," *ACM Transactions on Management Information Systems (TMIS)* (12:1), pp. 1-20.
Zhang, N. A., Wang, C. A., Karahanna, E., and Xu, Y. 2022. "Peer Privacy Concern: Conceptualization and Measurement," *MIS Quarterly* (46:1).

7. Appendix

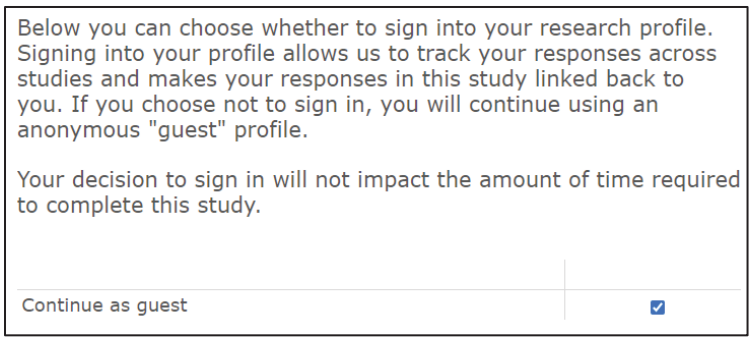
Appendix A: Experimental Design for Experiment 1 and 2



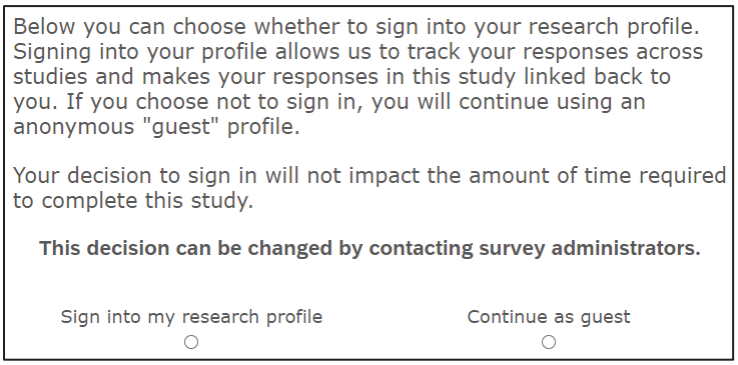
Dark Default Presentation



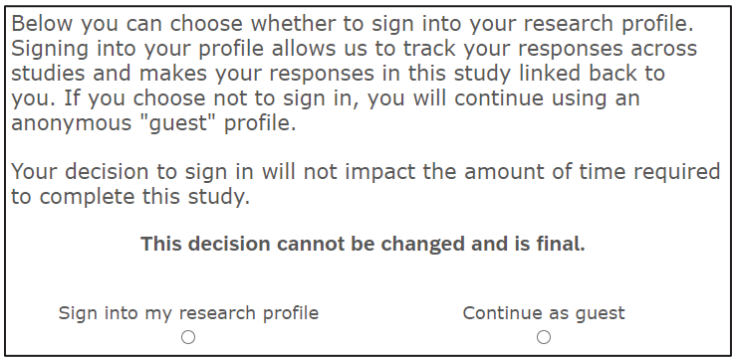
Active Choice Presentation



Protective Default Presentation



Reversible Presentation (with Active Choice Structure)



Irreversible Presentation (with Active Choice Structure)

Appendix B: Repeatability Effects for Experiment 1

VARIABLES	(Study 1) Log-In	(Study 2) Log-In	(Study 3) Log-In	(Panel) Log-In
AC	-0.114***	-0.120***	-0.118***	-0.115***
	(0.0259)	(0.0259)	(0.0264)	(0.0272)
PD	-0.446***	-0.463***	-0.460***	-0.441***
	(0.0315)	(0.0312)	(0.0315)	(0.0339)
Study (1-3)				-0.00289
				(0.00390)
AC × Study				-0.00157
				(0.00586)
PD × Study				-0.00795
				(0.00837)
Constant	0.903***	0.899***	0.819***	0.879***
	(0.0644)	(0.0596)	(0.0642)	(0.0540)
Observations	1,028	1,028	1,028	3,084
Number of ID				1,028

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix C: Disclosures for Experiment 1 and 2

Each question had the following possible answers: Yes, No, Prefer Not to Answer

Question (Experiment 1)	**Study**
While in a relationship, have you ever flirted with somebody other than your partner?	Romance
Have you ever used a dating app for the sole purpose of engaging in sexual activity?	Romance
Have you ever encouraged someone to drink when you were trying to seduce them?	Romance
Have you ever cheated while in a relationship?	Romance
Have you ever refrained from dating someone because of their skin color?	Romance
Have you ever looked at pornographic material?	Sex
Have you ever had a one night stand?	Sex
Have you ever taken nude pictures of yourself or a partner?	Sex
Have you ever had sexual thoughts about a member of your same sex?	Sex
Have you ever showered with a partner?	Sex
Have you ever used drugs of any kind (e.g. weed, heroin, crack)?	Crime
Have you ever let a friend drive after you thought he or she had had too much to drink?	Crime
Have you ever made up a serious excuse, such as grave illness or death in the family, to get out of doing something?	Crime
Have you ever stolen anything worth more than $50?	Crime

Have you ever downloaded pirated content from the internet?	Crime

Question (Experiment 2)
Do you like dogs more than cats?
Do you prefer showers over baths?
Do you like reading fiction more than non-fiction?
Do you prefer working in the mornings more than at night?
Do you like watching movies more than TV shows?

Appendix D: Exit Questions for Experiment 1 and 2

Each question was answered using a Likert-type scale with the following responses: Strongly Agree, Agree, Neither Agree nor Disagree, Disagree.

Question
I was concerned about my personal privacy when completing this study.
My responess could be used in a way that may harm me.
My responses are valuable to the researchers.
Maintaining the privacy of one's personal information is very important.
I trust the researchers with my responses.
I was comfortable signing into my research profile.
I was less honest when signed into my profile.
My decision to log in during one survey impacted my decision to log in for subsequent surveys.
I thought more about my decision each time I was asked to log in.
I am logged into my account most times that I use an online web service (Google, Amazon, YouTube, etc.).

Appendix E: Balance Checks for Each Experiment (Pairwise Comparisons for Group Means)

Experiment 1

Variable	Dark Default vs. Active Choice	Dark Default vs. Protective Default	Protective Default vs. Active Choice
Age	0.453	0.666	0.751
Male	0.157	**0.014**	0.297
Black	0.838	0.642	0.796
Asian	0.549	0.511	0.218
Hispanic	0.508	0.452	0.926
Advanced Degree	0.903	0.365	0.439
High School	0.290	0.324	0.949
Bachelor's	0.814	0.664	0.510

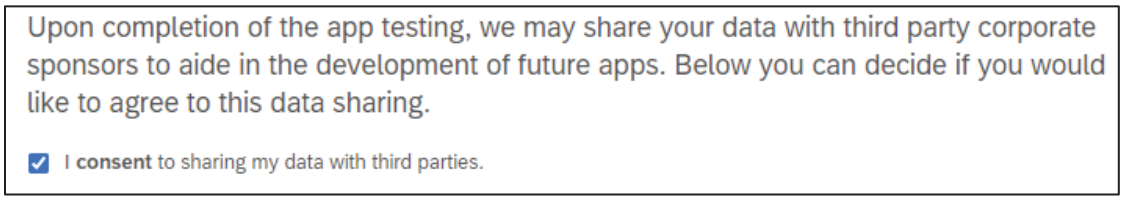
Experiment 2

Variable	Dark Default vs. Active Choice	Dark Default vs. Protective Default	Protective Default vs. Active Choice
Age	0.161	0.221	0.803
Male	**0.073**	0.641	0.180
Black	0.342	0.589	0.669
Asian	0.725	0.343	0.197
Hispanic	0.909	0.370	0.317
Advanced Degree	0.663	0.679	0.975
High School	0.312	0.690	0.532
Bachelor's	0.698	0.766	0.494

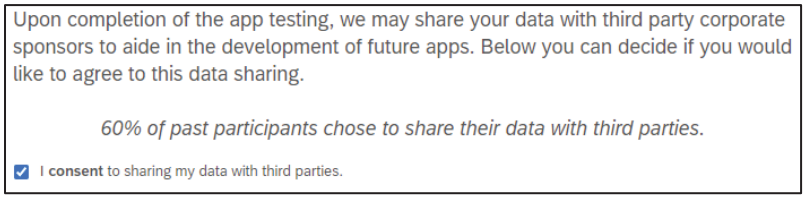
Experiment 3

Variable	Dark Default vs. Active Choice	Dark Default vs. Protective Default	Protective Default vs. Active Choice
Age	0.966	**0.086**	**0.087**
Male	0.579	0.850	0.453
Black	0.769	0.520	0.725
Asian	0.569	0.856	0.696
Hispanic	0.708	**0.049**	0.107
Advanced Degree	0.298	0.447	0.777
High School	0.366	0.814	0.499
Bachelor's	0.511	0.619	0.872

Appendix F: Examples of Consent Structure for Experiment 3

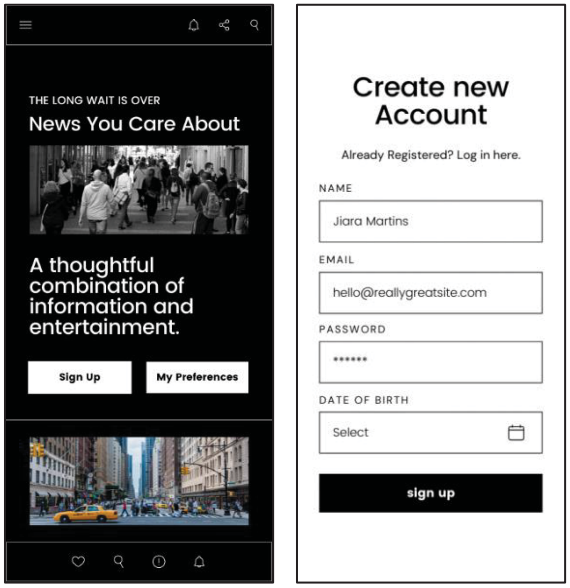


Dark Default Presentation

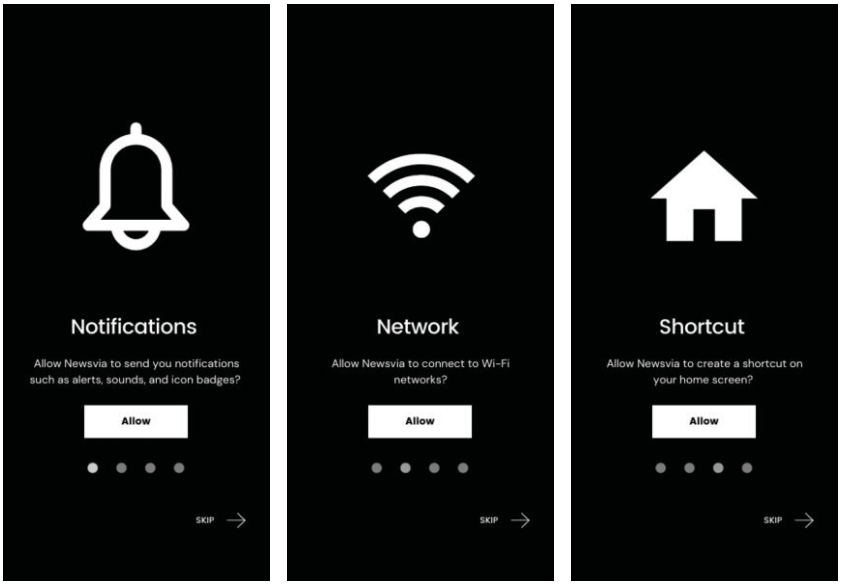


Social Norm Nudge

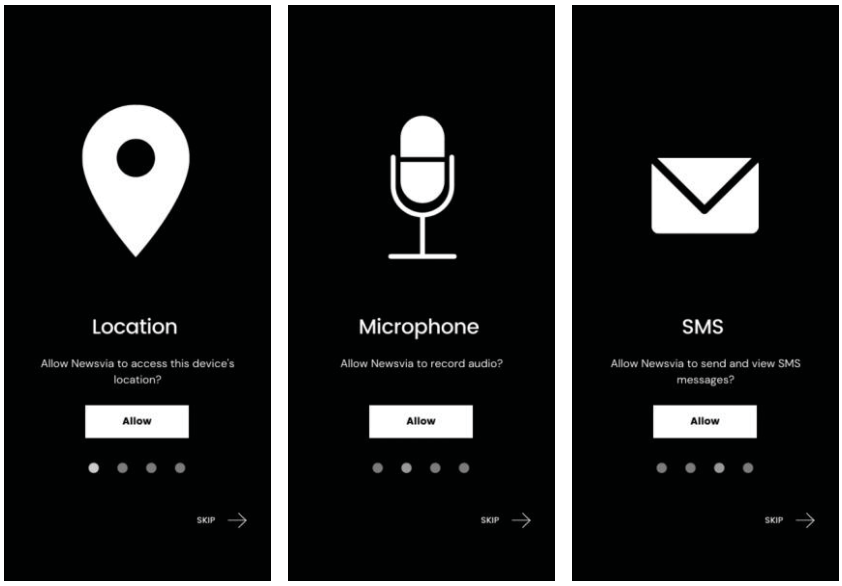
Appendix G: Application Renderings



Application Renderings in Experiment 3



Low-Risk Mobile App Permissions in Experiment 3



High-Risk Mobile App Permissions for Experiment 3

Appendix H: Survey Instrument for Experiment 3

Each question was answered using a Likert-scale with the following responses: Strongly Agree, Agree, Slightly Agree, Neither Agree nor Disagree, Slightly Disagree, Disagree, or Strongly Disagree.

Question	Context	Low-Risk Responses	High-Risk Responses
I am willing to download Newsvia.	Download Intention	5.3	3.3
After reading the related information about Newsvia, I am willing to try Newsvia.	Download Intention	5.5	3.5
After reading the related information about Newsvia, I am willing to consider Newsvia as a preferred app to download in the news and entertainment category.	Download Intention	5.0	3.4
I think Newsvia would be popular.	Perceived Popularity	5.0	4.2
The downloads of Newsvia will be numerous.	Perceived Popularity	5.1	4.0
I think Newsvia will be hot among users.	Perceived Popularity	4.5	4.1
The permissions requested by Newsvia are many.	Permissions Sensitivity	3.3	5.0
The personal information requested by Newsvia is sensitive.	Permissions Sensitivity	3.5	5.6

The potential risk related to the permission requests of the app is high.	Permissions Sensitivity	3.3	5.1
The permissions requested by Newsvia are appropriate for the needs of the app.	Permissions Sensitivity	5.1	3.1
The permissions requested by Newsvia are standard for applications like it.	Permissions Sensitivity	5.2	3.8
I would be willing to share my data from Newsvia with third-parties, if asked.	Permissions Sensitivity	3.6	3.0
I am sensitive to privacy-related issues.	General Privacy Concerns	5.5	5.8
To me, it is important to protect privacy.	General Privacy Concerns	5.7	6.2
I am concerned about potential privacy threats.	General Privacy Concerns	5.6	6.0

Chapter 2 – Till Tech Do Us Part: Betrayal Aversion and its Role in Algorithm Use

ABSTRACT

Failing to follow expert advice can have real and dangerous consequences. While any number of factors may lead a decision maker to refuse expert advice, the proliferation of algorithmic experts has further complexified the issue. One potential mechanism that restricts the acceptance of expert advice is betrayal aversion, or the strong dislike for the violation of trust norms. This study explores whether the introduction of expert algorithms in place of human experts can attenuate betrayal aversion and lead to higher overall rates of seeking expert advice. In other words, we ask: *are decision makers averse to algorithmic betrayal?* The answer to this question is uncertain ex ante; robust evidence exists showing that even inanimate products (e.g., airbags and vaccines) suffer reduced uptake due to betrayal aversion. We answer this question through an experimental financial market where there is an identical risk of betrayal from either a human or algorithmic financial advisor. We find that the willingness to delegate to human experts is significantly reduced by betrayal aversion, while no betrayal aversion is exhibited towards algorithmic experts. The impact of betrayal aversion towards financial advisors is considerable: the resulting unwillingness to take the advice of the human expert leads to a 20% decrease in subsequent earnings, while no loss in earnings is observed in the algorithmic expert condition. This study has significant implications for firms, policymakers, and consumers, specifically in the financial services industry.

1. **Introduction**

People often seek expert advice on investing for retirement, choosing a healthcare plan, and treating illness. However, individual uptake of expert advice has consistently been found to be suboptimal (Woodhouse and Nieusma 1997). In several contexts, including healthcare and financial planning, failing to adhere to expert advice can result in significant negative consequences (Seiders et al. 2015). Further increasing complexity, expert advice is rapidly being augmented by algorithms. Consumers are faced with the decision of whether to let an algorithm manage their pension investments (Lourenço et al. 2020), suggest potential dating partners (Prahl and Van Swol 2017), or control what digital content they read and watch (Prahl and Van Swol 2017; Van Swol 2011). This trend of algorithmic advice continues to increase, and likely will for years to come (Tetlock and Gardner 2016). Importantly, in many of these instances, evidence-based forecasts made by algorithms outperform those made by their human counterparts, making this new form of expert advice even more valuable (Beck et al. 2011; Dawes 1979; Highhouse 2008).

One potential mechanism that restricts the acceptance of expert advice is the fear of betrayal. Economics literature has shown that decision makers are often strongly averse to the possibility of a betrayal, even when controlling for monetary risk (Bohnet and Zeckhauser 2004). Take, for example, an expert financial advisor whose incentive structure may encourage them to occasionally make a self-interested financial decision: a betrayal. If an investor chooses to hire the advisor, that advisor could either 1) perform as expected or 2) betray the investor. The total utility for the investor would then consist of the expected *monetary* return and the *emotional* cost of being betrayed (betrayal aversion). An investor who, following a betrayal, experiences a positive emotional cost is betrayal averse and may choose not to hire the advisor to avoid experiencing the emotional disutility even if the monetary return is larger than any alternatives.

Our work aims to address what changes, if any, occur to levels of exhibited betrayal aversion when an expert is replaced by an algorithm. Critically, many accounts of betrayal aversion bypass the

need for human intentions completely and introduce a significant potential for betrayal aversion to persist for non-human advisors. For example, Koehler and Gershoff (2003)) show that safety products such as airbags or vaccines can elicit feelings of betrayal in their users. Other work shows that humans mindlessly attribute social rules and expectations to computers, even when they know the computers lack intentionality (Nass and Moon 2000; Nass et al. 1994). Currently, no work has examined if betrayal aversion persists when a human expert is augmented or replaced by an algorithm. This disconnect is essential to understanding the role of algorithms in motivating the acceptance of expert advice. Therefore, we ask: *are humans averse to algorithmic betrayal?*

Given the rapid growth of algorithm use in financial market trading[15] and its potential to increase efficiency, transparency, and capacity in the trading process (Bell and Gana 2012) , we contextualize our work within the financial services industry specifically. To answer our central research question, we conduct an economic lab experiment in which participants play a 40-round financial trading game using a simulated market structure and a real financial trading algorithm designed to give advice that would be comparable to what consumers would receive from both human and algorithmic experts in a real-life setting. Participants are given the choice in each round to either make their own decision on how much of their endowment to use to purchase a risky asset or use the advice of an expert. No deception was used in our experiment and all participants were compensated based on their actual investment decisions.

The experiment employs a two factor (2x3) between-subjects design. The first manipulated factor is the *presentation* of the expert as either human or algorithm; importantly, the underlying algorithm that generates the advice is identical for both the human and the algorithm expert. The second manipulated factor is the presence of a salient betrayal risk where treated participants are informed that there is a slight chance that the human or algorithmic expert may make a self-interested decision on their behalf. To isolate the emotional response to the betrayal (betrayal aversion), we include an error condition where

[15] https://therobusttrader.com/what-percentage-of-trading-is-algorithmic/

participants learn that there is a slight chance of an accidental error occurring. Ex ante, one would expect the error condition to elicit similar impacts on the perception of the monetary performance of the expert, while the lack of intentionality would limit the presence of betrayal aversion. We recruited 275 participants from Amazon Mechanical Turk (mTurk) to take part in the experiment over a video sharing platform in, what we believe to be, one of the first digitally-face-to-face economic experiments conducted on mTurk.

Our analysis captures changes in advisor usage trends in the betrayal risk (both human and algorithm), error risk, and control groups. When in a human advisor condition, we find a roughly 16% decrease (β=-0.159, p=0.024) in advisor usage when informed of a betrayal risk but no decrease when only informed of a risk of accidental error (β=0.007, p=0.924). This result implies that the effect we measure is due to betrayal aversion and not a perceived decrease in expected monetary return. Interestingly, we do not find this effect in the algorithm conditions (β=0.055, p=0.431). These findings imply that substituting an algorithmic advisor for a human advisor can significantly attenuate betrayal aversion. We also observe that betrayal aversion results in a significant economic loss, resulting in a loss of $2.15 (p=0.072), an almost 20% decrease in overall returns for affected participants. Analysis of exit questions substantiates that betrayal aversion may be a key aspect of this difference: participants in the betrayal condition with a human advisor reported feeling more concerned about being misled (p = 0.006), having their trust violated (p = 0.010), and most importantly, feeling betrayed (p = 0.001). We do not find these effects when the advisor was an algorithm, nor in the error-risk condition.

We find similar effects in a second experiment that utilizes the same experimental design while 1) drawing from a different population, undergraduate students recruited from the Virginia Tech Economics Lab and 2) utilizing a different human advisor. We find that risk of betrayal decreased uptake of the human advisor by 12% (p = 0.0901) and earnings by $2.13 (p = 0.003). The effects were again attenuated when assigned to an algorithm advisor (p = 0.026). This second experiment demonstrates the robustness of the findings to research platform, participant sample, and advisor selection. Note that the Virginia Tech

Economics labs explicitly prohibits any form of deception in experimental procedures, so this experiment reduces concerns about the credibility of the information provided to participants when compared to the mTurk sample.

Our work contributes to the nascent literature exploring the acceptance, or lack thereof, of expert algorithms. The majority of the literature (see Logg et al. (2019)) for a notable exception) highlights decision-maker's preference towards human experts (Dietvorst et al. 2015; Longoni et al. 2019) and identifies traits that algorithms *lack* that may lead an individual to exhibit "algorithm aversion" (Castelo et al. 2019). Proposed solutions to the nonacceptance of algorithmic experts have focused on either enhancing the human involvement in the algorithmic decision making by allowing users to add feedback (Dietvorst et al. 2018), making an algorithm more human-like by highlighting the affective abilities of the tool (Castelo et al. 2019) or anthropomorphizing the algorithms (Schanke et al. 2021). Our results highlight an important nuance to the conclusions of prior work: reducing the human element associated with algorithmic experts can remove some of the traditional barriers to the acceptance of expert advice.

We also contribute to this literature by identifying a potential strength of algorithmic experts and introducing a novel phenomenon to the algorithmic adoption literature – betrayal aversion. This phenomenon, largely unstudied by the information systems discipline, has been shown by behavioral and experimental economists to significantly influence an individual's decision-making. Given that the economics literature has shown the phenomenon to exist even when the betrayal comes from inanimate products or objects, its role in the human-algorithm relationship is essential to explore. However, no work, in economics or information systems, has explored the extent to which betrayal aversion may persist (or not) when a human is replaced by an algorithm. By bridging these two literatures, we show that despite prior evidence showing the potential of inanimate objects to elicit betrayal aversion, algorithms may attenuate the phenomenon instead.

We also provide valuable insights to the Financial Technology (FinTech) literature. Algorithmic trading services have grown substantially and trading services augmented by artificial intelligence are

likely in the near future (Gomber et al. 2018). Recent estimates show that between 60 and 75% of the total trading volume in the U.S. involves some form of algorithmic trading, but more research is needed to understand the implications of this trend (Alt et al. 2018; Cao et al. 2020; Gomber et al. 2018; Hendershott et al. 2021; Kou 2019). We specifically respond to Hendershott et al. (2021)) by examining the role of algorithmic financial advice and providing a potential solution to a significant barrier to the acceptance of expert advice within financial services.

2. Conceptual Background

Our work examines two primary areas of the literature. First, we look to previous work on algorithm adoption and aversion that focuses on specific drivers of use and disuse to better understand predictors of algorithmic acceptance. Then, we explore the betrayal aversion literature in economics and introduce it as a previously unexamined factor in algorithm uptake. In the section that follows, we contextualize our work to financial technology and examine the theoretical expectations of the effect of betrayal aversion on the use of trading algorithms.

2.1 Algorithm Adoption & Aversion

Prior work considering the use and adoption of information technology has strived to consider user's social and emotional beliefs (Benbasat and Wang 2005; Qiu and Benbasat 2005; Venkatesh and Davis 2000). This has led to the perception of IT artifacts as "social actors" (Al-Natour and Benbasat 2009; Reeves and Nass 1996). In the Computers are Social Actors (CASA) paradigm, humans have the same social rules and expectations of technology that they have of other humans (Nass et al. 1994). This paradigm now has renewed importance as digital ecosystems built on powerful customization algorithms are influencing the daily lives of their users (Parker et al. 2017). Artificial intelligence and algorithmic decision makers are augmenting or replacing their human counterparts, even in domains like medicine (Jussupow et al. 2021), and online services are increasingly using algorithms to determine the content users see (Orlikowski and Scott 2015). Given the increased prevalence of algorithm-run services and platforms

(Ransbotham et al. 2018), research that considers the complexities of algorithmic use and adoption is of growing importance.

While algorithm adoption has grown, some research has shown that overall, people still prefer human decision makers over algorithmic decision makers (Diab et al. 2011; Eastwood et al. 2012). Dietvorst et al. (2015)) terms this *algorithm aversion*, when individuals avoid using algorithms, after they see them make an error (See Burton et al. (2020)) for a review).

Several potential drivers to this phenomenon have been introduced. First, false expectations lead to an individual's unwillingness to use and accept an algorithm. This may occur because individuals believe that human error is random whereas algorithmic error demonstrates that the entire *system* is flawed instead of a one-time error (Dietvorst et al. 2015; Dietvorst et al. 2018; Highhouse 2008). Alternatively, individuals may value forming a professional relationship with those from whom they seek advice (Alexander et al. 2018; Önkal et al. 2009; Prahl and Van Swol 2017). Further, Castelo et al. (2019)) finds that algorithm aversion is dependent on task type, for example, when a task is perceived as subjective, individuals are less inclined to trust an algorithm. However, when individuals perceive that an algorithm is able to learn from experience, the effect of algorithm aversion is lessened (Berger et al. 2021). Additionally, when an individual gains experience with an algorithm, and receives feedback outlining the superior performance of the algorithm, aversion decreases over time (Filiz et al. 2021).

Research also shows that human decision makers fear relinquishing control, thus preferring a certain level of power over an algorithm's advice (Colarelli and Thompson 2008; Scherer et al. 2015). Dietvorst et al. (2018)) further explores the integration of human-in-the-loop decision making where the human decision maker oversees the algorithm's processes, creating a perception of control.

There is also a small but growing literature considering the opposite effect: that people prefer algorithms to humans. Some work has shown that when individuals are asked to remember important information, they opt to outsource the task to algorithms, showing a preference towards algorithms over

their own abilities (Sparrow et al. 2011). Further, Logg et al. (2019)) shows that individuals rely more heavily on algorithmic advice than advice from others or their own judgement, a phenomenon they term "algorithm *appreciation*." While most of the past work has proposed solutions to this debate by focusing on making algorithms more *human* (Schanke et al. 2021; Wilson et al. 2017), no work, to our knowledge, has considered the potential for algorithms to bypass traditional barriers to the willingness to seek advice often faced by human experts. We contribute to this literature by using insights from behavioral economics to explore the role of betrayal aversion in algorithm adoption.

2.2 Betrayal Aversion

Betrayal aversion is defined as the strong dislike for violations of trust norms implicit in a relationship between two parties (Aimone et al. 2015). Importantly, past work has shown that the degree of betrayal aversion depends on the extent to which the betrayers have a duty to protect. In Koehler and Gershoff (2003)), they show that participants view a betrayal as more severe when a member of the military commits treason than when its committed by an individual in an unrelated career. The same phenomenon can be seen with criminal punishment where, for example, people believe that day care workers who abuse a child should be punished more harshly than a janitor who abuses a child in the same way. This has led to demands to revise sentencing guidelines to make them sensitive to the extent to which the defendant's role included ensuring an individual's wellbeing (Shnoor 2009).

While many considerations of betrayal aversion have focused on human-to-human interactions, other work has acknowledged that non-human agents elicit fears of betrayal in the same way that humans do. Koehler and Gershoff (2003)) show this phenomenon through the elicitation of opinions on different vehicle air bags. They give participants the option to choose between Air Bag A that carries a 2% chance of driver death when involved in a serious accident or Air Bag B that carries a 1% chance of death in a serious car accident plus a 0.01% chance of the air bag causing the death when the driver would have otherwise survived. From a purely risk-minimizing perspective, Air Bag B is preferred. However, only 32.6% of participants in the experiment chose it. Similar findings from the same paper are produced when

considering smoke alarms and vaccines. Other work has shown that consumers who feel betrayed by products punish the firm by way of complaints or negative reviews (Grégoire and Fisher 2008).

Importantly, past work has focused on removing elements of risk and trust to isolate betrayal aversion as the primary explanatory variable. The general method of studying betrayal aversion through this lens involves having one group of participants play a trust game with random chance determining the player's financial return, instead of another human player. This work shows that participants require a higher minimum acceptable probability (MAP) of earning a return in a trust game played with a human compared to a similar game involving random chance, providing evidence of non-monetary betrayal aversion (Bohnet et al. 2008; Bohnet et al. 2010; Bohnet and Zeckhauser 2004; Hong and Bohnet 2007).

This phenomenon has expanded beyond economic decision making to other fields as well. Some work has shown that managers of are influenced by betrayal aversion when developing relationships with their employees (Birnberg and Zhang 2010). The results showed that some managers spend more money to prevent betrayal than they could have conceivably lost from the betrayal itself, implying that betrayal *aversion*, not just the betrayal, may lead to a decrease in expected monetary return. As firms increasingly are incorporating algorithmic tools and services into their underlying business models, the presence of betrayal aversion in both consumers and employees is likely to influence subsequent adoption and use.

3. Theoretical Development

Betrayal aversion with respect to algorithms has not yet been explored. To lend concreteness to our theoretical expectations, we focus on algorithm aversion in the financial services context. As mentioned previously, algorithmic trading (AT) has grown substantially in financial markets around the world. The use of AT has the ability to increase efficiency, transparency, and capacity in the trading process (Bell and Gana 2012). These increases are realized through the rapid collection of public information (Chakrabarty et al. 2015) and the ease in converting that information into prices (Brogaard et

al. 2014). Additionally, AT can enhance liquidity and increase the informativeness of price quotes (Hendershott et al. 2011).

Although investors and traders have increasingly developed trading algorithms designed to mimic the trends and actions of human traders (Hendershott et al. 2021), consumer unwillingness to adopt often superior algorithms can result in lower returns (Ge et al. 2021). In addition, if betrayal aversion persists in this context, the phenomenon should be considered in the design and marketing of algorithmic tools. While we defer to future work for specific design recommendations, we highlight for the first time the potential of algorithmic betrayal aversion and open the door for further explorations. In the rest of this section, we use the context of trading algorithms to explore the role of betrayal aversion in algorithmic use and leverage common assumptions of utility maximizing agents (e.g. Bernard et al. (2015))). What follows is not a formal analytical model but rather a conceptual exercise to help organize competing dynamics and articulate ex ante expectations around algorithmic betrayal aversion.

Assume that a decision maker has a basic utility function such that:

$$U(E, \sigma) = E - \frac{1}{2} R\sigma^2 \tag{1}$$

where E is the expected return or mean outcome of a financial market trade that follows a well-defined probability distribution function with a variance of σ^2. R is a constant risk aversion parameter that reflects the decision maker's risk preference (reflective of the Arrow-Pratt risk aversion measure first introduced by Pratt (1964)) and Arrow (1971))). Assume that the decision maker has a choice between two financial strategies for the trading decision. They can choose to make their own decision (S), or they can choose to use the recommendation of a financial advisor (A) such that:

$$U(E_i, \sigma_i) = E_i - \tfrac{1}{2} R\sigma_i^2, \text{ where } i \in [A, S] \tag{2}$$

Before making the choice, they are told the following:

$$E_A > E_S \tag{3}$$

Therefore, an individual would choose to use the advisor's recommendation over their own decision if the increase in expected return is higher than the difference in the variance scaled by the risk aversion parameter such that:

$$E_A - E_S > \frac{1}{2} R(\sigma_A^2 - \sigma_S^2) \tag{4}$$

However, they are also told that if A is chosen, there is a small probability of event B, or a betrayal, also occurring. This probability can be expressed as E_B (the mean probability of a betrayal occurring). The betrayal occurs when the advisor makes a self-interested choice that may or may not align with the ideal trading strategy of the decision maker, with equal probability. In other words, the betrayal may increase or decrease the investor's return. However, it is classified as a betrayal because the advisor makes the choice that best suits their own interests with no regards for the investor. If the goals of the advisor and the investor happen to align when a betrayal occurs, the investor may still financially benefit[16].

Therefore, the betrayal does not decrease their expected monetary return such that $E_{A|B=1} = E_{A|B=0}$. However, given the extensive empirical findings described in the previous section, we posit that the utility function also includes a non-monetary element β, or betrayal aversion, in which an individual exhibits an emotional cost from a betrayal that is distinct from any consideration of expected monetary returns. The difference in utility of choosing A over S would then be:

$$\Delta U = \left[E_A - \frac{1}{2} R\sigma_A^2 - \beta E_B\right] - \left[E_S - \frac{1}{2} R\sigma_S^2\right] \tag{5}$$

where β represents the level of betrayal aversion the individual exhibits and E_B is the expected betrayal outcome [0,1] as described above.

[16] While a real-world betrayal may come with financial consequences, for the sake of simplicity, we assume here that the expected return is constant. This allows us to isolate the betrayal aversion parameter more easily, increasing the clarity of our expectations.

Let us assume that there is an individual with a unique risk aversion parameter, R^*, who is indifferent between the two trading strategies assuming $E_A > E_S$ and $\sigma_A^2 > \sigma_S^2$:

$$E_A - \frac{1}{2}R^*\sigma_A^2 = E_S - \frac{1}{2}R^*\sigma_S^2 \tag{6}$$

$$R^* = \frac{2[E_A - E_S]}{\sigma_A^2 - \sigma_S^2} \tag{7}$$

If this individual also exhibits some level of betrayal aversion and the probability of a betrayal is non-zero, such that $\beta > 0; E_B > 0$, they will choose to make their own trading decision over taking the advisor's, as shown below:

$$E_A - \frac{1}{2}R^*\sigma_A^2 - \beta E_B < E_S - \frac{1}{2}R^*\sigma_S^2 \tag{8}$$

$$-\beta E_B < (E_S - E_A) + \frac{1}{2}\frac{2[E_A - E_S]}{\sigma_A^2 - \sigma_S^2}(\sigma_A^2 - \sigma_S^2) \tag{9}$$

$$\beta E_B > 0 \tag{10}$$

Therefore, in line with the central proposition of Bohnet and Zeckhauser (2004)), we hypothesize:

H1: Holding the objective monetary return constant, the willingness to outsource a decision to a human expert will decrease when there is a chance of betrayal.

This hypothesis is consistent with previous research on betrayal aversion towards a human counterpart. Since previous work has not examined betrayal aversion within financial investing, we first aim to explore whether betrayal aversion exists in this context. We then examine whether levels of betrayal aversion differ when a human advisor is replaced with an algorithmic advisor. Prior work has shown that humans mindlessly apply social rules and expectations to computers, even when they know that the computers lack feelings and intentionality (Nass and Moon 2000; Nass et al. 1994). Therefore, we anticipate betrayal aversion persisting for algorithmic experts. In other words, our expectations imply that:

$$\beta_{Algorithm} > 0 \tag{11}$$

leading us to hypothesize:

H2: Holding the objective monetary return constant, the willingness to outsource a decision to an algorithmic expert will decrease when there is a chance of betrayal.

Prior work has shown that the intensity of an individual's emotional response to betrayal depends on both the significance and depth of the relationship between parties and the magnitude of the harm caused by the betrayal (Rachman 2010). Revisiting our theoretical model, we can then further divide betrayal aversion, β into the magnitude of harm, H, and the depth of the relationship D, such that:

$$\beta = H + D \tag{12}$$

If one controls for the magnitude of harm (the decrease in expected return, in our case, H=0), the variance, and the risk preference between human and algorithm contexts, then it would follow that betrayal aversion would rely on the perceived significance of the decision-maker's relationship with the algorithm. Algorithms, generally, lack the ability to display social intelligence, and therefore the relationship between humans and algorithms is limited (Frey and Osborne 2017; Rafaeli et al. 2016). If this detracts from the perceived significance of the relationship a human feels they have with an algorithmic expert, we can assume that:

$$D_{Human} > D_{Algorithm} \tag{13}$$

Further showing that:

$$\beta_{Human} > \beta_{Algorithm} \tag{14}$$

Therefore, we hypothesize:

H3: Holding the objective monetary return constant, there will be a smaller decrease in the willingness to outsource a decision to an algorithm compared to a human when there is a chance of betrayal.

4. Financial Investment Game

We utilize a two-factor (2x3) between-subjects design. Our two factors differentiate the presentation of the advisor (human or algorithm) and the additional risk of some disutility (betrayal risk, accidental error risk, or no additional risk). The experimental task is a 40-round financial investing game, consisting of 10 baseline rounds (where the additional risk factor was not introduced) and 30 principal rounds. Each participant was assigned to either the human advisor or algorithm advisor treatment for the duration of the experiment. Participants start the task with a trading endowment which they can invest in a risky financial asset in our experimental market. The experiment is incentivized so that the cumulative return from 10 baseline rounds and 30 principal rounds determine a participant's earnings.

In both the human and algorithm advisor treatments, participants begin each round by deciding whether to choose their own level of investment or utilize an expert advisor. In the initial 10 baseline rounds there was no possibility of betrayal. At the start of the 30 principal rounds which followed, participants in the betrayal condition read a disclaimer indicating that there was incentive misalignment with the human/algorithm advisor that could result in occasional negative returns. Participants also learned that, historically, the human/algorithm advisor had outperformed participants' investment decision, thus ensuring the expected return for the advisor/algorithm would be higher than the expected return of investing themselves. For control, both the human and algorithm advisors offered the same advice, and we verified that the decision rule used by both advisor types (discussed below) truly did outperform participants' investment decisions. We included a shortened version of the disclaimer during each round to ensure the saliency of the treatment. The full disclaimer can be found in Appendix A.

4.1.1 Market Structure & Algorithm Design

The risky asset in our experimental market follows a two-state (good or bad) Markov-switching Gaussian random walk with a state switching probability of 35%[17], adapted from Zhang (2020)). In other words, if the previous round was a good state, there is a 35% chance that the following round would be a bad state and a 65% that it would be a good state. The expected return of the good state is described by:

$$r_t = \mu_1 dt + \sigma_1 dZ_t \tag{15}$$

where $\mu_1 = 0.10$, $\sigma_1 = 0.10$, and Z_t is white noise. The expected return of the bad state is described by:

$$r_t = \mu_2 dt + \sigma_2 dZ_t \tag{16}$$

where $\mu_2 = -0.05$ and $\sigma_2 = 0.05$. If the asset price increased, participants received the highest profit if they use their entire endowment to purchase the asset. If the asset price decreased, participants received the lowest loss by purchasing none of the asset. This allows us to account for the fictive error occurring in the market process. We can define the fictive error as:

$$f^+ = (100\% * r_t^+) - (Allocation * r_t^+), f^- = (0\% * r_t^-) - (Allocation * r_t^-) \tag{17}$$

where r^+ is a positive return on the asset and r^- is a negative return on the asset, $Allocation$ is the percentage of the participant's endowment that they used to purchase the risky asset, $Allocation * r_t^+$ is the experienced gain for the positive return, and $Allocation * r_t^-$ is the experienced loss for the negative return.

In both the human and algorithm conditions, the suggestion from the advisor was derived using the same Bayesian investment model. This allowed for a well-defined probability that the market was in a good state. Assume that r_t is a price change of the risky asset that the participant observed in period t. We then know that the probability that the market is currently in a good state is:

$$q_t = \Pr(s_t = good | r_t, r_{t-1}, \dots, r_2, r_1) \tag{18}$$

Therefore, we know:

[17] Simulations were run to identify an optimal switch rate that balanced the predictability of market fluctuations with the desire for the algorithm to consistently outperform the individual investors.

$$q_t(q_{t-1}, r_t) \tag{19}$$

$$= \frac{\Pr(r_t|s_t = good)\Pr(s_t = good|q_{t-1})}{\Pr(r_t|s_t = good)\Pr(s_t = good|q_{t-1}) + \Pr(r_t|s_t = bad)\Pr(s_t = bad|q_{t-1})}$$

Further, we know the expected return in a good state from (10). We can define the distribution function of the expected return $\Pr(r_t|s_t = good)$ as $f_{good}(r_t) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(r_t-\mu_1)^2}{2\sigma_1^2}}$. Similarly, in a bad state, the distribution function would be $f_{bad}(r_t) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(r_t+\mu_2)^2}{2\sigma_2^2}}$. Given the updated belief that the previous trial was in a good state q_{t-1}, $\Pr(s_t = good|q_{t-1}) = (1-p)q_{t-1} + p(1-q_{t-1})$ and $\Pr(s_t = bad|q_{t-1}) = pq_{t-1} + (1-p)(1-q_{t-1})$ where p is the state switch rate described above ($p = 0.35$). Therefore, we can rewrite (14) as following:

$$q_t(q_{t-1}, r_t) \tag{20}$$

$$= \frac{f_{good}(r_t)((1-p)q_{t-1} + p(1-q_{t-1}))}{f_{good}(r_t)\big((1-p)q_{t-1} + p(1-q_t)\big) + f_{bad}(r_t)(pq_{t-1} + (1-p)(1-q_{t-1}))}$$

The algorithm then, derives $q_t(q_{t-1}, r_t)$ based on the previous round's return to determine the probability of this round being in a good state. We then derive the optimal amount of the endowment to use using the maximization of the utility function subject to the relationships between expected return and standard deviation of the risky asset as follows:

$$y^* = \frac{q_t\mu_1 + (1-q_t)\mu_2}{0.02A\sigma^2} \tag{21}$$

where A is a measure of risk tolerance for an individual. To simplify the experiment, rather than measure individual risk preferences in advance we used the average risk aversion level (A=5.17) from Holt and Laury (2002)) shown in (21). We piloted 10 rounds of the trading task with a group of volunteers and had them make their own investment decisions in each round and recorded their investment returns. Afterwards, we simulated returns data using the rule used by the human/algorithm advisors with the same stream of market returns. We found that the returns earned by the advisor were 65% higher on average

than that earned by the participants, verifying that we were accurate in informing participants that utilizing the advisor would yield higher investment earnings was accurate.

4.1.2 Participants & Procedure

We recruited 460 participants through Amazon Mechanical Turk. A total of 275 participants attended their designated sessions and completed the experiment. We initially anticipated a 50% attendance rate given hesitation concerning the requirement to join Zoom and the safeguards put in place to filter out bots. Additionally, we restricted sign-ups to those located in the US along with minimum requirements for the number of HITs completed and the acceptance rates of those HITs. We advertised our study as a financial game with the opportunity to make on average \$20-\$30/hour[18] for participation.

Participants signed up for a designated time slot using a Qualtrics survey. They received email reminders 2 days prior to their time slot and 1 hour before the experiment was scheduled to begin. Emails were sent via automated scripts utilizing the mTurk API. Experiment sessions took place over a 2-week period. Even though this is an individual task, for efficiency, multiple participants completed the experiment at once. Each session began with participants joining a Zoom meeting room where the experimenter gave instructions. In the human condition, a third-year MBA student was introduced using the following statement: "This is Brandon. He is trained in the financial task you are being asked to complete today and will be aiding in the execution of the experiment." The student came to each session dressed professionally and the same human advisor was used for the duration of data collection to ensure no effects of advisor appearance or other characteristics. The procedure was the same in the algorithm advisor treatment, except that no advisor was introduced.

[18] The average AMT requester offers \$11/hour for work. We wanted our earnings potential to be particularly attractive to workers given the fact that most workers may not have experience with face-to-face HITs over Zoom and we wanted to complete data collection in a short period of time. The average earnings were \$10.29 + a \$5 show-up fee and the average completion time was 45 minutes. Therefore, our average hourly pay was \$20.34.

Participants were next sent a link that directed them to the online experiment. The experiment interface, including both the financial market and the investment algorithm, was developed using oTree Version 5.9 (Chen et al. 2016) and was deployed via Heroku, a commercial cloud hosting provider. Upon clicking the link, they were directed to the experiment and began by going through a brief tutorial that explained the market process and the task that they were being asked to complete. Then they were given the opportunity to invest in 10 unpaid practice rounds where they had to make their own investment decision in each round. After completing the practice rounds, participants were directed to a short comprehension quiz (see Appendix B). They were informed that they were being granted an endowment of 1000 research points with which to play the game, which translates to $5USD. They were also compensated an additional $5 which was added to their earnings at the end of the experiment.

At the onset of the 40 payoff determining rounds, participants were informed that the following rounds would be the same as the practice rounds with two key differences. First, these rounds would determine their true payoff at the end of the experiment. Second, they learned that they have the option in each round to make their own investment decision or to utilize a financial advisor/algorithm to make the decision for them. Before making their decision, they were directed to a waiting page where they saw a short video of the available advisor, either the human advisor or a graphic that visualized an algorithm working and were informed that the advisor was generating their investment recommendation based on market data (see Appendix C). This was to ensure that timing between rounds remained consistent and participants behavior was not impacted by perceptions of the length of time required to decide oneself relative to the advisor. It also improved the realism of the task since the advisor took a non-trivial amount of time to generate the recommendation. Importantly, prior to the first principal round, participants had no knowledge of the betrayal or risk treatment and were only told that they had the choice to use an advisor's (either an algorithm or a human) recommendation and that those who chose to use the advisor's recommendation earned more on average.

To avoid using deception, our human financial advisor took part in the actual experiment beyond being present in the Zoom session. His interface (shown in Appendix D) would report the suggested investment, using the same rule as the algorithm, for each participant in each round and would ask him to submit the recommendation. We kept the design as simple as possible to avoid any possibility of human error while avoiding deception. While participants may have believed that the advisor was manually developing investment recommendations, they were only ever told that the advisor would be submitting recommendations.

After the 10 baseline rounds participants in an assigned betrayal condition received a disclaimer that there was a small chance that the advisor or algorithm would intentionally over-invest even if he/it was not confident that it was a good market. Likewise, if the participant was in an error-risk condition, they were told that there was a small chance of an accidental error occurring. Again, the full disclaimers can be found in Appendix A.

Figure 1 shows the decision page for the control group. In the betrayal treatment and error treatment, this decision was accompanied by a small additional disclaimer stating, "The algorithm will occasionally over-invest even when it is not clear that it is a good market" and "The algorithm will occasionally make an accidental error", respectively, to ensure the saliency of our treatment. The decision of whether to use the advisor was made at the beginning of each round.

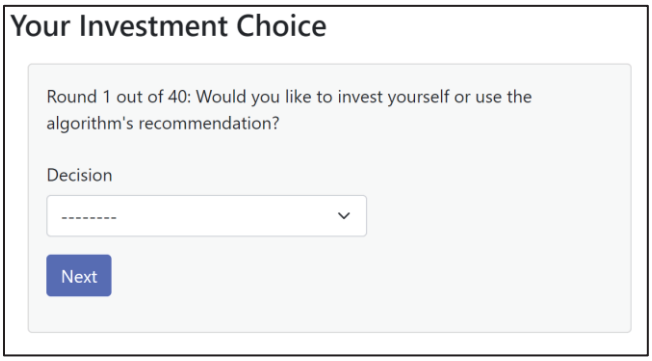


Figure 1: Decision Page

If the participant chose to invest themselves, they were directed to the investment page (Figure 2). Participants chose what percent of their current research points they wanted to use to purchase the risky asset. They were next directed to the results page (Figure 3) which displayed how much they invested, the percentage change in asset price for that round, their current round and cumulative returns (in research points), their total research points, as well as two graphs: one that tracks the price movement of the asset and the other that tracks their research points. If a participant chose to use the advisor's suggestion, they bypassed the investment page and went directly to the results.

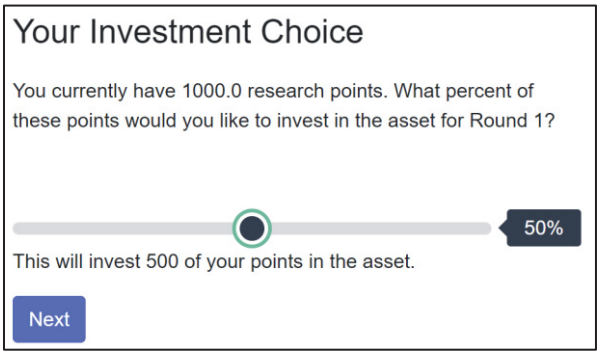


Figure 2: Investment Page

Figure 3: Results Page

After completing each of the investment decisions, participants answered a short demographic survey that asked for their gender, race, ethnicity, education, employment, experience with investing, and

familiarity with algorithm design. They also answered exit questions to measure their perceptions of trust, betrayal, and regret (see Appendix E). Participants earned $5 compensation in addition to the USD equivalent of their ending research points where 200 points equaled one dollar.

4.2 Estimation Approach

Our main dependent variable for our analyses is the participant's choice to use the advisor's recommendation or make their own decision in each round. Given the multi-round repeated structure of our experiment, we use a panel random effects model for our estimation.

$$\mathrm{Use}_{it} = \beta_i \times \mathbf{Advisor_i} + \beta_2 \times \mathbf{Risk_i} + \beta_3(\mathbf{Advisor_i} \times \mathbf{Risk_i}) + \alpha \times \mathbf{Demographics_i} + \delta_t + \theta_i + u_i \quad (22)$$

The dependent variable, Use_{it}, is a binary indicator that equals 1 if participant i chose to use the advisor in round t and 0 otherwise. $\mathrm{Advisor}_i$ is a vector of binary indicators for the advisor treatment condition assigned to each participant [algorithm vs. human]. Risk_i is a vector of binary indicators for the risk treatment condition assigned to each participant [control vs. error risk vs. betrayal risk]. We also include interactions between advisor presentation and risk treatments. $\mathrm{Demographics}_i$ is a vector of controls capturing heterogeneity in individual demographics (e.g., gender, race, ethnicity, education, work status, experience with algorithms, experience with investing). δ_t includes round fixed effects and θ_i is the participant-specific random effect. The error term, u_i, is clustered on participant. Estimates on the randomly assigned treatments are assumed unbiased due to the lack of correlation with unobserved individual differences and the error term. This assumption is tested in Section 5.1. This estimation approach allows for a robust account of time trends in our data and corrects for the nonindependence of multiple observations of the decision to use the advisor from a single participant.

Additionally, we estimate treatment effects for total returns. Since the algorithm, on average, realizes higher returns than participant decisions, any decrease in advisor usage should be accompanied by a decrease in returns as well.

5. Results

Two hundred and seventy-five individuals took part in our financial trading game. Table 1 shows the distribution of participants across treatments in our experiment. We find a roughly equal split for both the additional risk dimension (control vs. betrayal risk vs. error risk) and the advisor presentation dimension (algorithm vs. human). The average advisor usage was 56.44% for our baseline rounds and 42.59% in the 30 principal rounds.

	Algorithm	*Human*
Control	47	42
Error	51	42
Betrayal	46	47

Table 1: Sample Size by Experimental Conditions

Table 2 provides a description of our variables and summary statistics for each. We have a diverse sample of participants with a relatively even split between males and females and racial and ethnic statistics that are comparable to the national averages[19]. Fifteen participants indicated that they had extensive investment experience and only 7 indicated that they had extensive experience with algorithms.

Variable	Description	Mean (1)	S.D. (2)
Use	The choice to use the advisor in a specific round	0.461	0.498
Earnings	The total earnings from the game	9.141	5.292
Male	Whether the individual is a male	0.455	0.498
Caucasian	Whether the individual is Caucasian	0.749	0.434
African American	Whether the individual is African American	0.0945	0.293
Asian	Whether the individual is Asian	0.124	0.329
Hispanic	Whether the individual is Hispanic	0.0655	0.247
Some College	Completed some college	0.0982	0.298
Bachelor's Degree	Completed bachelor's degree	0.702	0.457
Advanced Degree	Completed advanced degree	0.200	0.400
Student	Whether the individual is a student	0.0473	0.212
Unemployed	Whether the individual is unemployed	0.135	0.341
Full-Time Employed	Whether the individual is employed full-time	0.0255	0.158
Retired	Whether the individual is retired	0.578	0.494
†Investment Experience	The participant's experience with investing	1.735	0.551

[19] https://www.census.gov/quickfacts/fact/table/US/PST045221

†Algorithm Experience	The participant's experience with algorithms	1.498	0.549

† Likert-type scale (1 = No Experience, 3 = Extensive Experience)
Table 2: Summary Statistics

5.1 Balance Checks

Before estimating our main effects, we evaluate the efficacy of random assignment by examining whether we there are differences in demographic characteristics across treatment groups. Table 3 provides pair-wise comparisons for all the variables listed in Table 2 across both the advisor presentation factor (Algorithm vs. Human) and the additional risk factor (Control vs. Betrayal vs. Error). We conduct 52 pairwise comparisons (13 variables * 4 comparisons) and find that 48 of these comparisons identify insignificant differences between conditions. With an alpha of 0.1, we would expect that 5 comparisons are significant by random chance. We identify only 4 significant differences between these groups. Even so, we control for these variables in our analysis and continue to identify consistent results.

Variable	Control vs. Betrayal	Control vs. Error	Betrayal vs. Error	Algorithm vs. Human
Male	0.663	0.109	0.238	0.180
Caucasian	0.302	0.882	0.233	0.810
African American	0.152	0.901	0.118	0.327
Asian	0.566	0.600	0.268	0.943
Hispanic	0.087*	0.341	0.422	0.780
Some College	0.408	0.609	0.176	0.981
Bachelor's	0.027**	0.140	0.459	0.472
Advanced Degree	0.320	0.243	0.861	0.952
Student	0.132	0.701	0.250	0.913
Full-Time Employed	0.109	0.367	0.011**	0.362
Other Employment	0.322	0.271	0.034**	0.328
Investment Experience	0.989	0.152	0.143	0.298
Algorithm Experience	0.900	0.693	0.583	0.412

Table 3: Comparison of Demographics between Conditions

Next, we leverage our first 10 baseline rounds which precede the introduction of any betrayal or error risk. We evaluate whether any differences exist across our risk conditions in use of the expert advisor and total earnings prior to the introduction of the additional risk. Table 4 estimates the baseline

effect of the experimental risk conditions. Additionally, Table 4 reports the baseline effect of the advisor presentation (human vs. algorithm) for the first 10 rounds.

We do not find any significant differences between risk conditions during the baseline rounds. For example, our primary advisor usage variable sees no significant difference between the betrayal risk treatment and our control ($p = 0.878$), the betrayal risk treatment and the error risk treatment ($p = 0.866$), or the error risk and control ($p = 0.996$). Interestingly, we also see no significant effects on advisor use between the algorithm and human conditions ($p = 0.770$) in these baseline rounds. Those in the algorithm treatment chose to use the advisor, on average, 56.9% of the time while those in the human treatment used the advisor 55.6%. Estimating treatment effects on total earnings for the first 10 rounds similarly provides no significant results. Overall, we find substantial balance across individual characteristics and outcomes prior to the introduction of betrayal.

VARIABLES	(1) Use	(2) Earnings	(3) Use	(4) Earnings
Betrayal Risk	0.00736	0.998		
	(0.0482)	(1.663)		
Error Risk	-0.000203	2.722		
	(0.0447)	(1.900)		
Algorithm			0.0110	-1.017
			(0.0376)	(1.512)
Constant	0.889***	18.75***	0.888***	20.51***
	(0.127)	(5.146)	(0.125)	(5.191)
Fixed Effects	YES	NO	YES	NO
Demographics	YES	YES	YES	YES
Observations	2,750	275	2,750	275
Number of ID	275	275	275	275

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Baseline Effects – Ten Rounds Prior to Risk Treatment

5.2 Main Effects

Absent the presence of betrayal or error, we find no significant difference in uptake between the human and algorithmic advisor ($p=0.566$). However, this trend changes when the potential for betrayal is introduced. Figure 4 graphs the average advisor usage across each of the treatment sets. First, we find a

significant drop in human advisor usage for the betrayal risk treatment (Control μ=45.1%; Error μ=45.2%; Betrayal μ=29.1%). However, we see no such drop off in advisor usage for the algorithm treatments (Control μ=40.6%; Error μ=48.9%; Betrayal μ=46.2%). The lack of a negative effect in the error conditions implies that the betrayal effect for the human advisor is due to betrayal aversion and not a perceived decrease in expected return or increase in risk generally. Further, this provides preliminary evidence that substituting a human expert with an algorithm may attenuate betrayal aversion.

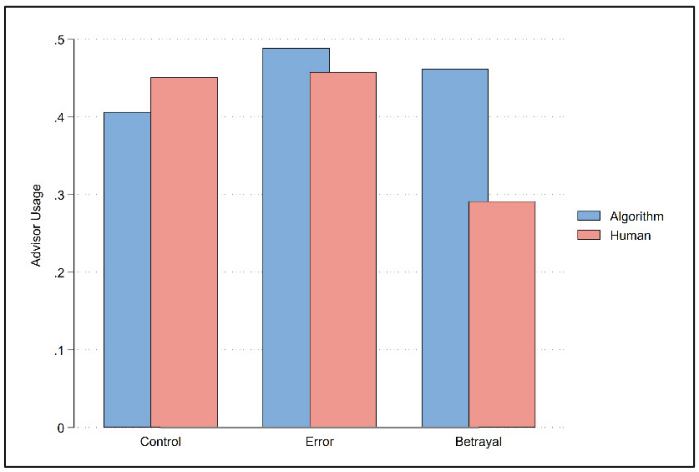


Figure 4: Advisor Usage Across Conditions

Our regression model confirms our summary results. The estimates are reported in Table 5. First, we find no significant effect of the algorithm presentation, absent additional betrayal or error risk (p=0.563) (column (1)). On average, our estimates show that in the human conditions, adding betrayal risk decreases use by roughly 16% (p=0.023) (column (1)). Conversely, introducing risk of an accidental error has no effect on usage (p=0.924) (column (1)). This implies that betrayal aversion is observed for the human conditions. Importantly, we find a positive and significant interaction effect of roughly +21% between betrayal risk and the algorithm treatment (p=0.031) (column (1)). This implies that the betrayal aversion found in the human treatments is *largely* attenuated when the human advisor is switched out for an algorithm. These effects are consistent when controlling for time fixed effects and participant demographics (columns (2) and (3)).

	Overall			1-10	11-20	21-30
	(1)	(2)	(3)	(4)	(5)	(6)

VARIABLES	Use	Use	Use	Use	Use	Use
Betrayal Risk	-0.159**	-0.159**	-0.155**	-0.147**	-0.163**	-0.161*
	(0.0702)	(0.0704)	(0.0706)	(0.0708)	(0.0772)	(0.0823)
Error Risk	0.00714	0.00714	0.00841	0.0310	0.0466	-0.0492
	(0.0753)	(0.0754)	(0.0756)	(0.0745)	(0.0864)	(0.0866)
Algorithm	-0.0444	-0.0444	-0.0284	-0.0106	-0.0156	-0.0651
	(0.0768)	(0.0770)	(0.0781)	(0.0770)	(0.0859)	(0.0885)
Algorithm × Betrayal	0.215**	0.215**	0.178*	0.141	0.206*	0.214*
	(0.0992)	(0.0994)	(0.101)	(0.101)	(0.110)	(0.116)
Algorithm × Error	0.0754	0.0754	0.0680	0.00835	0.0544	0.140
	(0.103)	(0.103)	(0.101)	(0.104)	(0.115)	(0.117)
Constant	0.451***	0.421***	0.600***	0.547***	0.618***	0.709***
	(0.0563)	(0.0602)	(0.146)	(0.0598)	(0.167)	(0.166)
Fixed Effects	NO	YES	YES	YES	YES	YES
Demographics	NO	NO	YES	YES	YES	YES
Observations	8,250	8,250	8,250	2,750	2,750	2,750
Number of ID	275	275	275	275	275	275

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 5: Main Effect of Treatment on Use

5.3.1 Persistence of Treatment Effects

We also explore heterogeneity across time for our treatment effects. Figure 5a and 5b plot average advisor usage across rounds for each condition. In Figure 5a, we can clearly see the effect of betrayal for the human conditions. It appears as if the effect remains relatively constant across rounds. In Figure 5b, algorithm usage rates for the betrayal treatment appear to drop initially before rising by the tenth principal round. This could imply that the algorithm experienced some initial betrayal effects but, unlike the human conditions, those effects dissipated quickly.

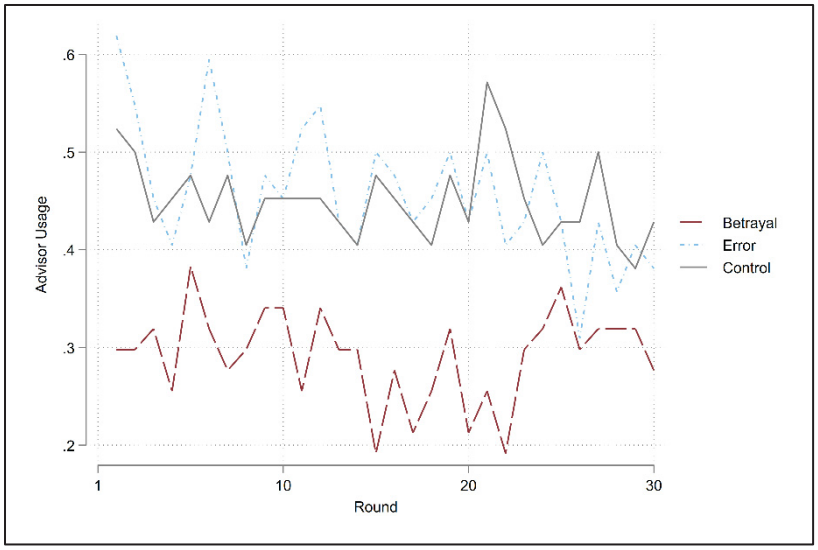


Figure 5a: Human Usage Rates Across Treatments

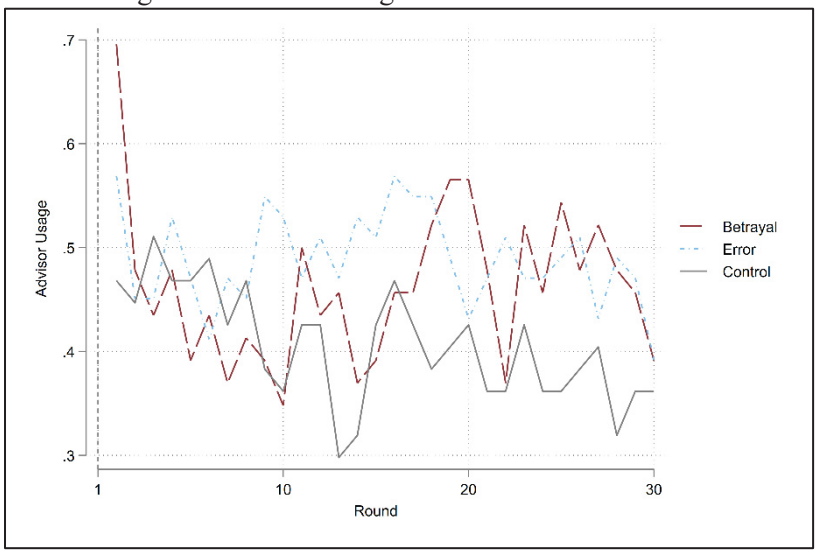


Figure 5b: Algorithm Usage Rates Across Treatments

Table 5 above confirms these results. We estimate our model for the first 10 principal rounds (1-10) (column (4)), the next 10 principal rounds (11-20) (column (5)), and the last 10 principal rounds (21-30) (column (6)). While betrayal aversion towards the human advisor remains consistent across rounds, we find that in the first 10 treated rounds, substituting the human with an algorithm does not fully attenuate the betrayal aversion (p=0.266) (column (3)). However, after at least 10 treated rounds, the betrayal aversion towards algorithms is completely attenuated (p=0.058) (column (4)). As stated above, this provides evidence of some initial, but not persistent, betrayal aversion for algorithms.

5.3.2 Investment Returns

We estimate betrayal treatment effects for the total earnings received by each participant (Table 6). Given that the human advisor uses the algorithm to make decisions and there is no difference in expected return for the betrayal or error conditions, if a participant chose the advisor over investing themselves in any condition, we would expect the same average returns. However, given the lack of advisor usage in the human betrayal condition, we analyze the effect of this decreased uptake on subsequent earnings. Our results indicate a $2.15 decrease in total earnings for those in betrayal treatment with a human advisor (p=0.072). We observe no significant effects of the error conditions for either advisor (Human p=0.195; Algorithm p=0.281) nor is there an effect of betrayal risk for the algorithm treatment (p=0.883). This supports the conclusion that the decrease in earnings is due to participants choosing to utilize the advisor less in the human betrayal condition. Given that the average earnings for those in the control condition with a human advisor was $10.29 (not including the $5 show up fee), a decrease of $2.15 would equate to a 21% drop in total earnings. These findings highlight the real financial consequences of betrayal aversion.

	Human	Algorithm
	(1)	(2)
VARIABLES	Earnings	Earnings
Betrayal Risk	-2.153*	0.180
	(1.186)	(1.223)
Error Risk	-1.502	-1.118
	(1.153)	(1.032)
Constant	10.29***	9.577***
	(0.933)	(0.835)
Observations	131	144
R-squared	0.030	0.012

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 6: Betrayal Effect on Total Earnings

5.3.3 Exit Questions

In Table 7 we estimate the effect of our treatments on several exit questions related to feelings of being misled, having trust violated, and importantly, feeling betrayed. The questions were answered with

a 5-point Likert-type scale where 1 = strongly disagree and 5 = strongly agree. Positive coefficients imply a stronger trend towards agreeing with the question. Column 1 reports estimates for responses to the question, "If you chose to invest with the advisor, you were concerned about being misled." For the human condition, betrayal has a positive and statistically significant effect (p=0.006). This effect is almost entirely cancelled out by the algorithm and betrayal interaction term (p=0.048) implying that feelings of being misled from the betrayal risk were not a concern in the algorithm conditions. Column 2 reports estimates to the question asking if participants were concerned about their trust being violated. Again, we find similar results to the first question with the caveat that the algorithm interaction term is slightly insignificant (p=0.103).

VARIABLES	(1) Misled	(2) Trust Violated	(3) Feel Betrayed
Algorithm	0.245	0.0755	0.390
	(0.242)	(0.236)	(0.250)
Betrayal Risk	0.671***	0.650**	0.815***
	(0.240)	(0.252)	(0.254)
Error Risk	0.0476	-0.0476	0.214
	(0.256)	(0.249)	(0.264)
Algorithm × Betrayal	-0.663**	-0.568	-1.095***
	(0.333)	(0.347)	(0.357)
Algorithm × Error	0.0825	0.164	-0.185
	(0.339)	(0.338)	(0.358)
Constant	3.095***	2.690***	2.738***
	(0.183)	(0.175)	(0.177)
Observations	275	275	275
R-squared	0.038	0.038	0.046

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 7: Exit Question Analysis

Our primary question of concern asked participants if they had concerns about feeling betrayed by the advisor. This was our main mechanism check given that the treatment does not explicitly use the word "betrayal" but implies a betrayal in the form of an incentive misalignment. The effect of the treatment in Column 3 is the largest coefficient (p=0.001). This implies that our treatment worked as intended and elicited feelings of betrayal in participants. Additionally, the interaction term for the

algorithm condition cancels out the treatment effect entirely (p=0.002), again providing evidence that algorithms can attenuate betrayal aversion.

6. Robustness: Experiment 2

To validate the results of our first experiment, we conducted a condensed version of the investment game with a new sample: undergraduate students. Unlike the main study, we do not include an additional error treatment (only a betrayal risk treatment and the control). Finally, there is a within-subjects fee treatment that is introduced after the first 20 rounds which decreased advisor use to almost zero across conditions. Therefore, we chose to analyze only the first 20 rounds where there were no fees. All other procedures are identical to the main study.

There are several key benefits to this follow-up study. First, we utilize the Virginia Tech Economics lab to recruit our sample. Researchers that recruit through this lab are prohibited from using any form of deception in their experiments. This is explicitly highlighted to participants, and they should therefore understand that any claims made throughout the experiment are true. This additional confidence in the veracity of information provided by the researchers help alleviate concerns around the believability of the information provided to participants in the first experiment. Additionally, we utilized a different human financial advisor to further provide robustness around advisor appearance and helps alleviate concerns of effects specific to a particular individual.

One hundred and twenty-three participants took part in our experiment. Table 8 shows the results of our panel random effects models that we used to estimate our treatment effects. Our effects from the main study are replicated here. Column 1 estimates a roughly 12% decrease in advisor usage when betrayal risk is introduced ($p = 0.091$). Conversely, we see no significant decrease in advisor usage when betrayal risk is introduced to the algorithm conditions ($p = 0.167$). Column 3 reports the interaction term between betrayal risk and the algorithm, confirming that switching the advisor from a human to an algorithm counteracts the measured betrayal aversion ($p = 0.026$).

Likewise, Column 4 shows that betrayal aversion decreased total earnings by roughly $2.13 (p = 0.003), strikingly similar to the decrease found in the main study. Those in the algorithm conditions saw no such decrease in earnings (p = 0.387, column (5)). The results from Experiment 2 confirm findings from Experiment 1 across the board. This level of robustness is difficult to attain even when replicating experimental procedures that draw from the same population. Provided that Experiment 2 drew from an entirely new population with large demographic differences (e.g., college-aged students versus a far older population), this is an impressive replication that provides significant richness to our findings.

	Human	Algorithm	All	Human	Algorithm
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Use	Use	Use	Earnings	Earnings
Betrayal Risk	-0.123*	0.0774	-0.123*		
	(0.0712)	(0.0561)	(0.0708)		
Algorithm			-0.118*	-2.128***	-0.566
			(0.0697)	(0.709)	(0.651)
Algorithm × Betrayal			0.201**		
			(0.0902)		
Constant	0.423***	0.305***	0.423***	18.72***	17.62***
	(0.0560)	(0.0421)	(0.0557)	(0.593)	(0.555)
Number of ID	48	75	123	48	75

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Table 8: Betrayal Effect on Advisor Use and Earnings (Experiment 2)

7. Discussion: Implications for Research and Practice

Our results show that while human experts elicit significant betrayal aversion, substituting those experts with algorithms may attenuate most, if not all, of the effect. Additionally, we highlight the real financial consequences of betrayal aversion and show a 20% decrease in earnings due to betrayal. Finally, we show that while algorithms may experience some initial betrayal aversion, it quickly subsides, whereas the aversion persists for human experts.

This work is not without limitations. First, we only focus on the context of financial investing. Future work could explore this avenue by studying contexts with higher stakes and exploring the effect that stakes have on betrayal aversion and algorithm aversion. Moreover, determining if our findings are

consistent across numerous settings would provide further insights. Additionally, the experiment is not able to measure individual level betrayal due to the between-subjects design. Aimone et al. (2015)) propose a betrayal aversion elicitation task that uses a within-subjects design to determine betrayal aversion at the individual level. Future research could adopt this framework to uncover heterogeneity around individual characteristics and their impact on betrayal aversion. Further, this paper only considers the role of betrayal aversion as it realtes to algorithms. In reality, the choice to use an algorithm is likely influenced by several other factors including familiarity with the algorithm provider, transparency of the algorithm, autonomy over the algorithm, and others. Future research could incorporate these ideas and explore their initial effects and interactive effects with betrayal aversion on algorithm adoption and use.

Despite these limitations, our work has significant implications for research. Our findings highlight the value in looking to behavioral and experimental economics to uncover potential attributes of algorithic tools that drive an individual to utilize an expert advisor. Further, we show that studying the adoption of algorithms and the interactions that humans have with them should not be viewed as an isolated phenomenon. Instead, by comparing individual behavior with other humans versus an algorithm, we can better understand what mechanisms drive adoption. For example, one possible explanation for the attenuation of betrayal averison by algorithms is that participants in the human condition were able to interact with their advisor. This element of social connection heightens trust, which subsequently increases concerns over betrayal. Researchers can use these findings to develop solutions to hestitation around accepting algorithmic tools that utilize the inherent stregnths of the algorithm itself. Further, future research could extend this potential pathway by varying the levels of social connection that an individual has with an algorithm.

Finally, our findings highlight the need for behavioral economcis to revisit previously well-accepted phenomena in light of advancing technologies. While the majority view of the literature is that inanimate objects can elicit betrayal aversion, we provide evidence to the contrary. One potential reason for this deviation could be that past work considering inanimate objects and betrayal aversion focused

exclusively on products meant to protect consumer's *physical* safety. It may be the case that while algorithms can attenuate betrayal aversion in a financial services setting, they may still elicit the fear when a consumer's physical outcomes are at risk. Future work can extend this thought by looking for instances of betrayal aversion with smart cities and Internet-of-Things devices that direct traffic or deploy emergency services, betrayal aversion elicited through online health portals, or the effect of algorithmic vaccine deployment on betrayal fears. Each of these pathways would prove fruitful in developing a deeper understanding of human decision-making and thought.

Industry could also benefit from the findings we present. Largely, firms across industries are adopting algorithmic solutions and determining which roles within their corporations could be augmented by algorithms. While efficiencies and improved operations play a large role in this consideration, acceptance from shareholders and customers is equally important. Our work provides valuable insight into the relationship humans have with algorithms and how that relationship leads to adoption and acceptance. For example, firms that provide face-to-face services that rely on consumer acceptance of their output may benefit from enhanced algorithmic participation, given the decrease in social expectations of algorithms from the consumer.

More specifically, financial services can utilize our results to improve uptake of financial planning services. While firms are increasingly introducing algorithmic tools to industry customers, there is little work showing the value of these tools to individual consumers. By augmenting traditional financier involvement in strategic customer planning with advanced algorithmic tools, consumers may be more willing to seek expert advice.

Overall, our results highlight the potential of algorithms to decrease the trust burden that many consumers face with "experts" that they encounter. Mitigating betrayal aversion can lead to increased demand for expert advice and better individual outcomes for consumers. The potential of algorithms from the perspective of efficiency and accuracy is only one piece of the larger puzzle. Algorithms may make

individuals feel at ease and allow them to overlook some of the less impactful social rules that might have restricted their previous gains.

8. References

Aimone, J., Ball, S., and King-Casas, B. 2015. "The Betrayal Aversion Elicitation Task: An Individual Level Betrayal Aversion Measure," *PLOS ONE* (10:9), p. e0137491.

Al-Natour, S., and Benbasat, I. 2009. "The Adoption and Use of It Artifacts: A New Interaction-Centric Model for the Study of User-Artifact Relationships," *J. AIS* (10).

Alexander, V., Blinder, C., and Zak, P. J. 2018. "Why Trust an Algorithm? Performance, Cognition, and Neurophysiology," *Computers in Human Behavior* (89), pp. 279-288.

Alt, R., Beck, R., and Smits, M. T. 2018. "Fintech and the Transformation of the Financial Industry." Springer, pp. 235-243.

Arrow, K. J. 1971. "The Theory of Risk Aversion," *Essays in the theory of risk-bearing*), pp. 90-120.

Beck, A. H., Sangoi, A. R., Leung, S., Marinelli, R. J., Nielsen, T. O., van de Vijver, M. J., West, R. B., van de Rijn, M., and Koller, D. 2011. "Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival," *Sci Transl Med* (3:108), p. 108ra113.

Bell, D., and Gana, L. 2012. "Algorithmic Trading Systems: A Multifaceted View of Adoption," *2012 45th Hawaii International Conference on System Sciences*: IEEE, pp. 3090-3099.

Benbasat, I., and Wang, W. 2005. "Trust in and Adoption of Online Recommendation Agents," *J. AIS* (6).

Berger, B., Adam, M., Rühr, A., and Benlian, A. 2021. "Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn," *Business & Information Systems Engineering* (63:1), pp. 55-68.

Bernard, C., Chen, J. S., and Vanduffel, S. 2015. "Rationalizing Investors' Choices," *Journal of Mathematical Economics* (59), pp. 10-23.

Birnberg, J., and Zhang, Y. 2010. "When Betrayal Aversion Meets Loss Aversion: The Effects of Changes in Economic Conditions on Internal Control System Choices," *Journal of Management Accounting Research* (23).

Bohnet, I., Greig, F., Herrmann, B., and Zeckhauser, R. 2008. "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States," *American Economic Review* (98:1), pp. 294-310.

Bohnet, I., Herrmann, B., and Zeckhauser, R. 2010. "Trust and the Reference Points for Trustworthiness in Gulf and Western Countries," *The Quarterly Journal of Economics* (125:2), pp. 811-828.

Bohnet, I., and Zeckhauser, R. 2004. "Trust, Risk and Betrayal," *Journal of Economic Behavior & Organization* (55:4), pp. 467-484.

Brogaard, J., Hendershott, T., and Riordan, R. 2014. "High-Frequency Trading and Price Discovery," *The Review of Financial Studies* (27:8), pp. 2267-2306.

Burton, J. W., Stein, M.-K., and Jensen, T. B. 2020. "A Systematic Review of Algorithm Aversion in Augmented Decision Making," *Journal of Behavioral Decision Making* (33:2), pp. 220-239.

Cao, L., Yuan, G., Leung, T., and Zhang, W. 2020. "Special Issue on Ai and Fintech: The Challenge Ahead," *IEEE Intelligent Systems* (35:2), pp. 3-6.
Castelo, N., Bos, M. W., and Lehmann, D. R. 2019. "Task-Dependent Algorithm Aversion," *Journal of Marketing Research* (56:5), pp. 809-825.
Chakrabarty, B., Moulton, P., and Wang, X. 2015. "Attention Effects in a High-Frequency World,").
Chen, D. L., Schonger, M., and Wickens, C. 2016. "Otree—an Open-Source Platform for Laboratory, Online, and Field Experiments," *Journal of Behavioral and Experimental Finance* (9), pp. 88-97.
Colarelli, S. M., and Thompson, M. 2008. "Stubborn Reliance on Human Nature in Employee Selection: Statistical Decision Aids Are Evolutionarily Novel," *Industrial and Organizational Psychology* (1:3), pp. 347-351.
Dawes, R. M. 1979. "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist* (34:7), pp. 571-582.
Diab, D., Pui, S.-Y., Yankelevich, M., and Highhouse, S. 2011. "Lay Perceptions of Selection Decision Aids in Us and Non-Us Samples," *International Journal of Selection and Assessment* (19).
Dietvorst, B. J., Simmons, J. P., and Massey, C. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err," *Journal of Experimental Psychology: General* (144:1), pp. 114-126.
Dietvorst, B. J., Simmons, J. P., and Massey, C. 2018. "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science* (64:3), pp. 1155-1170.
Eastwood, J., Snook, B., and Luther, K. 2012. "What People Want from Their Professionals: Attitudes toward Decision-Making Strategies," *Journal of Behavioral Decision Making* (25).
Filiz, I., Judek, J. R., Lorenz, M., and Spiwoks, M. 2021. "Reducing Algorithm Aversion through Experience," *Journal of Behavioral and Experimental Finance*), p. 100524.
Frey, C. B., and Osborne, M. A. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?," *Technological Forecasting and Social Change* (114), pp. 254-280.
Ge, R., Zheng, Z., Tian, X., and Liao, L. 2021. "Human–Robot Interaction: When Investors Adjust the Usage of Robo-Advisors in Peer-to-Peer Lending," *Information Systems Research*).
Gomber, P., Kauffman, R. J., Parker, C., and Weber, B. W. 2018. "Financial Information Systems and the Fintech Revolution." Taylor & Francis, pp. 12-18.
Grégoire, Y., and Fisher, R. J. 2008. "Customer Betrayal and Retaliation: When Your Best Customers Become Your Worst Enemies," *Journal of the Academy of Marketing Science* (36:2), pp. 247-261.
Hendershott, T., Jones, C. M., and Menkveld, A. J. 2011. "Does Algorithmic Trading Improve Liquidity?," *The Journal of Finance* (66:1), pp. 1-33.
Hendershott, T., Zhang, X., Zhao, J. L., and Zheng, Z. 2021. "Fintech as a Game Changer: Overview of Research Frontiers," *Information Systems Research* (32:1), pp. 1-17.
Highhouse, S. 2008. "Stubborn Reliance on Intuition and Subjectivity in Employee Selection," *Industrial and Organizational Psychology: Perspectives on Science and Practice* (1:3), pp. 333-342.

Holt, C. A., and Laury, S. K. 2002. "Risk Aversion and Incentive Effects," *The American Economic Review* (92:5), pp. 1644-1655.
Hong, K., and Bohnet, I. 2007. "Status and Distrust: The Relevance of Inequality and Betrayal Aversion," *Journal of Economic Psychology* (28:2), pp. 197-213.
Jussupow, E., Spohrer, K., Heinzl, A., and Gawlitza, J. 2021. "Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence," *Information Systems Research*).
Koehler, J. J., and Gershoff, A. D. 2003. "Betrayal Aversion: When Agents of Protection Become Agents of Harm," *Organizational Behavior and Human Decision Processes* (90:2), pp. 244-261.
Kou, G. 2019. "Introduction to the Special Issue on Fintech." SpringerOpen, pp. 1-3.
Logg, J. M., Minson, J. A., and Moore, D. A. 2019. "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes* (151), pp. 90-103.
Longoni, C., Bonezzi, A., and Morewedge, C. K. 2019. "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research* (46:4), pp. 629-650.
Lourenço, C. J. S., Dellaert, B. G. C., and Donkers, B. 2020. "Whose Algorithm Says So: The Relationships between Type of Firm, Perceptions of Trust and Expertise, and the Acceptance of Financial Robo-Advice," *Journal of Interactive Marketing* (49), pp. 107-124.
Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81-103.
Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers Are Social Actors," *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72-78.
Önkal, D., Goodwin, P., Thomson, M., GönüL, S., and Pollock, A. 2009. "The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments," *Journal of Behavioral Decision Making* (22:4), pp. 390-409.
Orlikowski, W. J., and Scott, S. V. 2015. "The Algorithm and the Crowd: Considering the Materiality of Service Innovation," *MIS Q.* (39), pp. 201-216.
Parker, G., Van Alstyne, M., and Jiang, X. 2017. "Platform Ecosystems:: How Deve/Opers Lnvert the Firm," *MIS Quarterly* (41:1), pp. 255-266.
Prahl, A., and Van Swol, L. 2017. "Understanding Algorithm Aversion: When Is Advice from Automation Discounted?," *Journal of Forecasting* (36:6), pp. 691-702.
Pratt, J. W. 1964. "Risk Aversion in the Small and in the Large," *Econometrica* (32:1/2), pp. 122-136.
Qiu, L., and Benbasat, I. 2005. "Online Consumer Trust and Live Help Interfaces: The Effects of Text-to-Speech Voice and Three-Dimensional Avatars," *International Journal of Human-Computer Interaction* (19:1), pp. 75-94.
Rachman, S. 2010. "Betrayal: A Psychological Analysis," *Behaviour Research and Therapy* (48:4), pp. 304-311.
Rafaeli, A., Altman, D., Gremler, D. D., Huang, M.-H., Grewal, D., Iyer, B., Parasuraman, A., and de Ruyter, K. 2016. "The Future of Frontline Research: Invited Commentaries," *Journal of Service Research* (20:1), pp. 91-99.
Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., and Spira, M. 2018. "Artificial Intelligence in Business Gets Real," MIT Sloan Management Review and The Boston Consulting Group.

Reeves, B., and Nass, C. I. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY, US: Cambridge University Press.
Schanke, S., Burtch, G., and Ray, G. 2021. "Estimating the Impact of "Humanizing" Customer Service Chatbots," *Information Systems Research* (32:3), pp. 736-751.
Scherer, L. D., de Vries, M., Zikmund-Fisher, B. J., Witteman, H. O., and Fagerlin, A. 2015. "Trust in Deliberation: The Consequences of Deliberative Decision Strategies for Medical Decisions," *Health Psychol* (34:11), pp. 1090-1099.
Seiders, K., Flynn, A. G., Berry, L. L., and Haws, K. L. 2015. "Motivating Customers to Adhere to Expert Advice in Professional Services: A Medical Service Context," *Journal of Service Research* (18:1), pp. 39-58.
Shnoor, B. 2009. "Loss of Chance: A Behavioral Analysis of the Difference between Medical Negligence and Toxic Torts,").
Sparrow, B., Liu, J., and Wegner Daniel, M. 2011. "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips," *Science* (333:6043), pp. 776-778.
Tetlock, P. E., and Gardner, D. 2016. *Superforecasting: The Art and Science of Prediction*. Random House.
Van Swol, L. M. 2011. "Forecasting Another's Enjoyment Versus Giving the Right Answer: Trust, Shared Values, Task Effects, and Confidence in Improving the Acceptance of Advice," *International Journal of Forecasting* (27:1), pp. 103-120.
Venkatesh, V., and Davis, F. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46), pp. 186-204.
Wilson, H., Daugherty, P., and Bianzino, N. 2017. "When Ai Becomes the New Face of Your Brand," *Harvard Business Review* (27).
Woodhouse, E. J., and Nieusma, D. 1997. "When Expert Advice Works, and When It Does Not," *IEEE Technology and Society Magazine* (16:1), pp. 23-29.
Zhang, X. 2020. "Experiential and Neurobiological Influences on Economic Preferences and Risky Decision Making." Virginia Tech.

9. Appendices

Appendix A: Risk Disclaimers

Disclaimer for Betrayal Treatments (Verbiage in brackets refers to the human condition):

"In the past 10 rounds, you were given the option to use a *financial investment algorithm* [financial advisor, Brandon] to make your investment decisions. Historically, *the algorithm* [Brandon] has outperformed those who choose to invest themselves, where the majority of negative returns *it* [he] received were the result of random market volatility. However, *investment algorithms, like the one used in this*

experiment, are designed to earn revenue [he is compensated] based on how frequently *it* [he] invests. Occasionally *the algorithm* [he] will intentionally over-invest even when *the algorithm* [he] is not confident that it is a good market. This incentive misalignment comes at the expense of increasing your risk and accounts for a small proportion of historical negative returns."

Disclaimer for Error Treatments (Verbiage in brackets refers to the human condition)

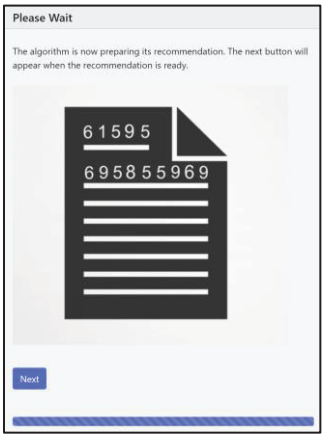
"However, *investment algorithms, like the one used in this experiment,* [his recommendations] are not always perfect and [he] has occasionally made accidental errors that result in a negative return. These errors come at the expense of increasing your risk and account for a small proportion of historical negative returns."

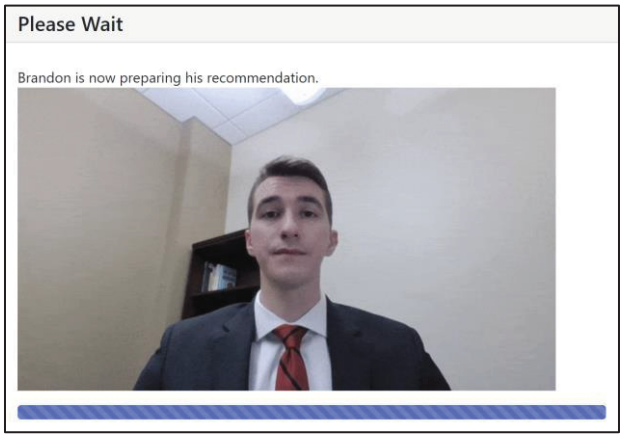
Appendix B: Comprehension Quiz

Each question had the following possible answers: True, False

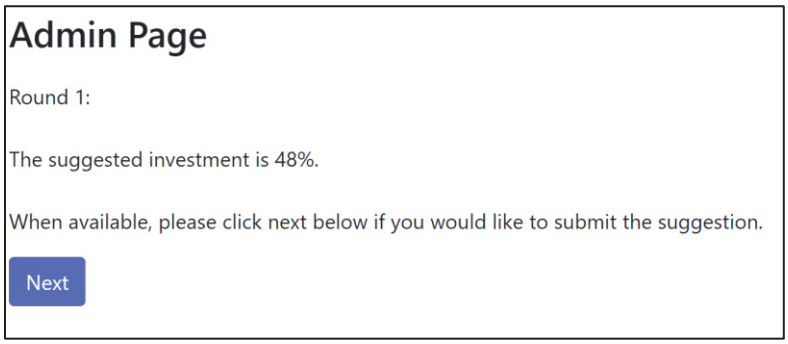
Question	**Answer**
You are being asked to invest your research points in a market with TWO possible assets.	False – there is only ONE asset.
You can invest 0% of your research points in the asset.	True.
If the CURRENT market is good, then there is a higher chance of the NEXT market being bad.	False – If the current market is good, there is a higher chance of the next market being good.
1000 research points translates to $1 at the end of the experiment.	False – 1000 research points translates to $5 at the end of the experiment.
If you have any questions throughout the experiment, you should ask them through Zoom only.	True.

Appendix C: Waiting Page (Left – Algorithm Treatment; Right – Human Treatment)





Appendix D: Advisor Interface



Appendix E: Exit Questions

Each question was answered using a Likert-type scale with the following responses: Strongly Agree, Agree, Neither Agree nor Disagree, Disagree.

Question
If I chose to invest with help, I was concerned about being misled.
If I chose to invest with help, I was concerned about my trust being violated.
If I chose to invest with help, I would feel betrayed if I received a negative return.
If I chose to invest with help, I trusted that a good investment decision would be made for me.
If I chose to invest with help, I felt like I made a mistake if I received a negative return.
If I chose to invest on my own, I felt like I made a mistake if I received a negative return.

Chapter 3 – Set Your Status to Away: Nomophobia and its Impact on Employee Well-Being

ABSTRACT

This paper explores the phenomenon of nomophobia, or the fear of being without a mobile phone, in the context of the workplace. With an increasing reliance on mobile phones for work-related tasks, there is a growing need to understand how this fear affects employees. The study aims to answer key research questions, including the prevalence of workplace nomophobia, the impact of moderators (such as age, race, personality type, and family status) on the severity of nomophobia, and the effects of nomophobia on productivity. The research utilizes a unique dataset collected from information workers across the United States, employing innovative approaches to estimate the prevalence of nomophobia and its impact on stress and productivity. This study introduces the concept of nomophobia to the information systems discipline, offers valuable insights for industry leaders, policymakers, and individuals, and contributes to the understanding of the evolving challenges presented by technology dependence in the modern workplace.

1. Introduction

The proliferation of mobile phones has largely shaped the technological landscape of the 21st century (Ling 2004). The nature of content generation (Ghose and Han 2011), knowledge gathering (Dellarocas et al. 2015), social capital (Ganju et al. 2016), commerce (Einav et al. 2014), and work (Chen and Karahanna 2018) have all transformed in response to this phenomenon. Questions around the merit of such a shift have been debated in both the press[20] and academia (Soror et al. 2015). Mobile phone use has been shown to improve social connectedness (Ran and Lo 2006), commerce for small and medium enterprises (Khaskheli et al. 2017), and information capture (Thakur et al. 2011). However, others have highlighted the potential pitfalls for phone use and overuse including diminishing mental and physical health (Elhai et al. 2017), lower productivity at work (Galluch et al. 2015), and even parental neglect (Mi et al. 2023).

Increasingly, dependence on mobile phones may sometimes lead to feelings of emotional distress when left without a device, referred to as nomophobia (King et al. 2013). Specifically, *nomophobia* is "the modern fear of being unable to communicate through a mobile phone" (Yildirim and Correia 2015). Concerningly, a recent meta-analysis that included 12,462 subjects identified 21% of individuals as suffering from severe nomophobia (Humood et al. 2021). Recent proposals have requested that nomophobia be included in future editions of the Diagnostic Statistical Manual of Mental Disorders (DSM) and by current definitions, it is considered a "specific phobia" diagnosable by DSM-IV (Bragazzi and Del Puente 2014). The implications of nomophobia are considerable, with those who have it reflecting suboptimal

[20] Have Smartphones Destroyed a Generation? *The Atlantic.* https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/

coping strategies for stress (Bragazzi et al. 2019), diminished academic performance (Aldhahir et al. 2023), and overall lower life satisfaction (Sharma et al. 2019)

As dependence on mobile phones continues to grow, these implications will only be exacerbated. One relevant context, currently understudied in the nomophobia literature, is the workplace (Wang et al. 2018). Several distinct features of the workplace make the consideration of nomophobia particularly relevant. First, employees are increasingly required to use mobile phones for work-related tasks. This gives rise to competing pressures between employee productivity and well-being (Magni et al. 2023). Whereas enhanced connectivity to a mobile device may lead to productivity gains within organizations (Sarker et al. 2010), it also may diminish employee mental health as feelings of being unable to escape from work are heightened (Butts et al. 2015). Further, the rapid shift to remote work, in response to the COVID-19 pandemic, has had a myriad of complex implications for employees (Galanti et al. 2021). Although working from home can have negative impacts on work-life balance and feelings of social isolation, a frequent concern of those who experience nomophobia is proximity to home (Yildirim and Correia 2015) making the ex-ante effects of remote work on nomophobia unclear. The following interview quote from Yildirim and Correia (2015) highlights this phenomenon:

> "If [my mobile phone] does go dead, that's the sort of thing when it is like 'I need to charge my phone right now.' Especially, if I'm not at home and it dies, it is just an uncertainty of like what if I forgot my keys? If it does die, you lose a peace of mind."

In this work, we seek to address the role of nomophobia in the workplace. We operationalize our phenomenon of interest by examining the relationship between phone battery level and stress. Phone-battery level is a suitable avenue for exploring nomophobia given the

direct relationship between battery levels and the perceived threat of losing access to a device. Specifically, we pose the following research questions: 1) what is the effect of phone-battery level on employee stress, 2) what is the indirect effect of phone-battery level on productivity that operates through stress, and 3) how does working from home moderate the relationship between phone-battery level-related stress and productivity? To answer these questions, we utilize a unique dataset collected on information workers across the United States.

The data we utilize was collected by The Tesserae Project (Mattingly et al. 2019); a large-scale field study conducted in 2018. The project utilized multimodal sensing, including location beacons, a wearable, a phone application, and daily surveys to track employee behavior and productivity for over 700 information workers across the United States. To develop our identification strategy for the causal effect of phone-battery level on stress, we develop a directed acyclic graph to identify potential confounders. Our resulting analyses utilize an instrumental variable approach to address endogeneity in our model. We then conduct moderated-mediation analysis to address our second and third research objectives.

This work has several important contributions. First, we introduce the concept of nomophobia to the information systems discipline. Given the field's substantial focus on mobile phone use and dependence, this unique perspective is a suitable fit for contextualizing future work. Further, past work on nomophobia has focused extensively on its impact in the personal lives of phone users but few studies have extended this consideration to the workplace. Those that have looked at workplace nomophobia have largely relied on cross-sectional data. The access to longitudinal data provided to us by the Tesserae Project allows for a more robust consideration of the main effects. Finally, by introducing remote work status as a moderator to

our main effects, we lend important insights to the nascent literature considering the rapidly changing landscape of work.

Outside of academia, our findings contribute to the ability of industry leaders to manage and care for the well-being of their employees, especially as cross-generational differences emerge. Further, analyzing productivity implications adds to the relevance of our findings for organizations. Policymakers may find our results helpful in guiding the enactment of regulations aimed at protecting employees and, our work may assist in providing further evidence in support of the significant mental health toll employees face from increased technology dependence. Finally, our findings may allow individuals further insight into their own relationship with technology and mobile devices specifically. Understanding how phone-related stress affects their daily lives may allow for a more nuanced self-management of mental health and overall well-being.

2. Conceptual Background & Hypotheses

To explore the role of nomophobia in the workplace, we first look at literature considering the behavioral implications of mobile phone use. Then we contextualize our consideration to the workplace and explore the impact of mobile phones on stress and productivity. Finally, we utilize literature that examines clinical nomophobia to bridge the two preceding domains.

2.1 Mobile Phone Use

Mobile phones have become an integral part of modern life, reshaping how people connect, communicate, and access information. The prevalence of these devices has given rise to a complex interplay between the utilization of mobile phones and well-being, reflecting several

positive and negative implications. On one hand, mobile phones may enhance social relationships and provide a sense of solidarity and camaraderie among users (Kwon et al. 2016). This constant connectivity allows for perpetual contact, fostering a sense of kinship. Conversely, overdependence on mobile phones and social applications can lead to feelings of inadequacy and potentially addictive behaviors with adverse psychological consequences (Griffiths 2005; Kwon et al. 2016).

Compulsive behaviors associated with mobile phone use, such as checking social media feeds incessantly, interrupting sleep for digital updates, and using phones in potentially dangerous situations (e.g., driving), highlight the issues this dependency may cause (Hedges 2014). This phenomenon is driven by various factors, including self-identity, a desire for belongingness, and dysfunctional coping mechanisms (Kim et al. 2011; Kuss and Griffiths 2011; Pelling and White 2009). There is also a universal appeal and potential risk of these devices. For example, prior works has shown that both extroverts and introverts are heavily drawn to social media, albeit for different reasons (Amichai-Hamburger and Vinitzky 2010).

A focus on technology *addiction* has been a dominant perspective in addressing this phenomenon (D'Arcy et al. 2014; Haug et al. 2015; Takao et al. 2009; Turel and Serenko 2010). However, contextualizing this relationship through the lens of constant checking and state-tracking behaviors, as introduced by Gerlach and Cenfetelli (2020), provides a more nuanced view. They propose that constantly checking your phone is driven by the need to stay updated and is facilitated by the easy accessibility of information through smartphones. This behavior is often cited as a significant source of stress, as the incessant need to stay connected may lead to heightened anxiety and distraction (Chang 2015; Condliffe 2017). This constant state of

connectivity, while enabling perpetual contact with social circles and the sharing of daily routines, can also lead to the feeling of being constantly on edge (Kwon et al. 2016).

Although constant connectivity may lead to IT-related overload and a myriad of cognitive and emotional symptoms (Maslach and Jackson 1981; Rutkowski and Saunders 2018), it can also serve practical and emotional needs. Interestingly, some work has shown that mobile phones can provide a sense of psychological comfort to their users (Melumad and Pham 2020). In moments of stress, users are often drawn to their smartphones to alleviate negative emotions. This dual role of mobile phones as both a stressor and a comforter highlights the complexity of their impact on subsequent human behavior.

A smaller subset of the literature has specifically examined the role of phone-battery level on its relationship to anxiety and stress. The anxiety surrounding decreasing phone-battery level is substantial, with 90% of individuals reporting feelings of panic when their phone's battery drops below 20%. This concern drives behaviors such as frequent charging and the alteration of usage patterns in response to changing battery levels (Ferreira et al. 2011; Hosio et al. 2016). When a device battery is low, individuals heighten their perceived value of the smartphone and become increasingly concerned about losing access to it (Hosio et al. 2016). This phenomenon is especially pronounced in situations where charging opportunities are scarce (e.g., during travel or at crowded events) (Hosio et al. 2016). A low battery symbolizes the potential loss of a lifeline to social networks, information, and entertainment, thus heightening users' anxiety. In contrast, when users are near charging facilities, most often their home, the perceived value of the battery diminishes, and the stress associated with battery life is reduced (Hosio et al. 2016).

Prior work has highlighted the multifaced relationship between phone use and stress. However, the substantial reliance on mobile phones coupled with heightened anxiety associated with decreasing battery levels is well-established. Users, already conditioned to rely on their phones for social and emotional stability, may experience heightened stress as their battery depletes, not just due to the potential disconnection from their social networks, but also due to the loss of a personal anchor and a source of psychological comfort. Therefore, we hypothesize:

H1: Decreasing phone-battery level increases stress.

2.1.1 Workplace Phone Use

Beyond the individual behavioral and well-being implications of mobile phone use, organizations are similarly tasked with a multi-faceted relationship between phones, productivity, and employee well-being. In the contemporary workplace, mobile phones have largely reshaped the relationship employees have with their work. Mobile phones provide unparalleled flexibility, allowing workers to perform tasks and communicate from virtually anywhere, thus enhancing their ability to manage work responsibilities (Derks and Bakker 2014; Mazmanian et al. 2013). This ubiquitous connectivity can lead to increased work engagement and facilitate knowledge management, especially in contexts where remote work has become the norm (Pandey et al. 2021; Zhai et al. 2023). Additionally, mobile phone use can aid in maintaining continuous communication with colleagues and clients (Santoro et al. 2018; Serenko et al. 2016).

However, this constant connectivity also blurs the boundaries between work and personal life (Chen and Karahanna 2018), potentially leading to significant work-life conflict (Sarker et al. 2012). Excessive use, especially during non-work hours, can lead to work-life conflict, emotional exhaustion, and reduced well-being (Boswell and Olson-Buchanan 2007; Király et al.

2020; Yu et al. 2018). The intrusion of work into personal time, facilitated by mobile phones, has often been associated with declines in productivity (Brunborg et al. 2011; Lanaj et al. 2014; Liu et al. 2021). Furthermore, the intense use of mobile phones for work purposes can lead to knowledge hiding and reduced empathy among employees, undermining the work environment (Choudhary and Mishra 2023; Zhang and Ji 2022).

One interesting area of the literature highlights the use of mobile phones for simultaneous evening work and leisure, illustrating the trade-off between productivity and well-being. Evening work-related use of mobile devices can lead to reduced sleep quantity, affecting next-day work engagement and potentially increasing work-related exhaustion (Lanaj et al. 2014). Conversely, evening leisure activities on these devices can offer psychological detachment from work, potentially improving sleep quality and, by extension, work-related outcomes (Hülsheger et al. 2015).

Reexamining the literature discussed in the preceding section on phone-battery life lends important insights to the workplace as well. Interestingly, the perceived value of mobile phones, which plays a crucial role in how they are used, is impacted substantially by their battery levels (Hosio et al. 2016). Users often adjust their usage patterns based on battery levels, conserving battery for more "valuable" uses as it depletes (Ferreira et al. 2011). As battery life decreases, employees may become more judicious in their use of mobile phones, potentially reducing distractions and improving focus on work tasks. This aligns with prior work motivated by the conservation of resources theory (Hobfoll 1989). Work productivity could be enhanced in light of the more efficient use of time and resources. Conversely, decreasing battery could also hinder productivity by limiting access to work-related communications and resources, thereby

increasing stress and anxiety related to inaccessibility and potential missed opportunities (Chen and Karahanna 2018; Tams et al. 2018).

Given the apparent dual effects of mobile phone use in the workplace, we propose the following competing hypotheses:

H2a: Phone-battery level has a positive direct effect on productivity.
H2b: Phone-battery level has a negative direct effect on productivity.

2.2 Nomophobia

Nomophobia, defined as the fear of being without one's mobile phone, represents a modern psychological condition intertwined with the ubiquity of mobile technology and its integral role in daily life. Coined during a 2008 study by the UK Post Office, nomophobia encapsulates the anxiety experienced when out of contact with a mobile device (King et al. 2010; SecurEnvoy 2012). The proliferation of smartphones, noted for their multifunctional capabilities, has significantly contributed to this phenomenon (Kang and Jung 2014; Park et al. 2013). Smartphones serve not just as communication tools but as essential devices for information access, social networking, and various daily tasks, leading to a heightened dependency on their constant availability (Park et al. 2013; Yildirim and Correia 2015).

The dimensions of nomophobia encompass fears related to the inability to communicate, loss of connectedness, restricted access to information, and the forfeiting of convenience offered by smartphones (Yildirim and Correia 2015). These dimensions highlight a deep-seated reliance on mobile technology for maintaining social interactions and accessing information in real-time. The anxiety associated with potential disconnection extends beyond mere discomfort but often manifests in significant stress (Yildirim and Correia 2015). Nomophobia-induced stress is

commonly associated with decreasing phone-battery level. The thought of a phone running out of battery can evoke significant distress, as it symbolically cuts off the user from their social networks, information streams, and sense of security. Users have reported their smartphone as a "peace of mind," associating a charged battery with freedom from stress and anxiety (Yildirim and Correia 2015).

Nomophobia's prevalence varies across demographic factors. Whereas the 2008 study revealed that over half of mobile phone users suffered from nomophobia, there was a higher incidence in men and it predominantly affected younger adults (DailyMail 2008). Further, the pathology appears to be particularly prevalent in student groups (Bartwal and Nath 2020; Farooqui et al. 2018) and associated with negative learning outcomes and diminished academic performance (Ahmed et al. 2019; Lee et al. 2018; Mendoza et al. 2018; Prasad et al. 2017). Across these groups, nomophobia is linked to mental disorders, self-esteem issues, loneliness, and overall happiness (Kuscu et al. 2021; Lee et al. 2018).

Extending these findings to the workplace, it is plausible that nomophobia could similarly decrease work productivity. The anxiety and stress induced by nomophobia, driven by social threats and the perceived need for constant connectivity and responsiveness (King et al. 2013; Tams et al. 2018), can potentially disrupt focus and decrease cognitive resources. Further, the dependence on mobile technology and the consequent stress from being disconnected can lead to constant distractions directly impeding organizational productivity (Aguilera-Manrique et al. 2018; Ayyagari et al. 2011; Samaha and Hawi 2016). The demand-control-person model postulates that such stress arises from threats to valued social resources (Rubino et al. 2012). This model suggests a direct link between social phobias like nomophobia and stress, where access to a mobile phone is the social resource being threatened when battery diminishes.

Although workplace-specific literature on nomophobia is under-studied, the implications of the pathology have clear links to workplace productivity. Therefore, we hypothesize that:

H3: Phone-battery level has a negative indirect effect on productivity mediated through stress.

Further drawing on the demand-control-person model, the role of remote-work may have important interactions with nomophobia and productivity. First, remote work could lessen the perceived social threat and the need for constant availability and immediate responsiveness, which are central to nomophobia-induced stress (King et al. 2013; Tams et al. 2018). Working from home offers more control and flexibility over one's environment, potentially mitigating the stress and anxiety associated with nomophobia (Galluch et al. 2015). Further, remote work environments offer a crucial advantage: constant and easy access to charging facilities. The proximity to personal charging solutions can significantly reduce the anxiety associated with a draining phone battery. This reduction in stress is not merely about ensuring a charged device but also about the psychological comfort derived from knowing that the means to recharge is within easy reach. Therefore, we hypothesize that:

H4: The indirect effect of phone-battery level on productivity mediated through stress is moderated by working from home.

Figure 1 below provides a visual representation of our conceptual model, based on the hypotheses given above.

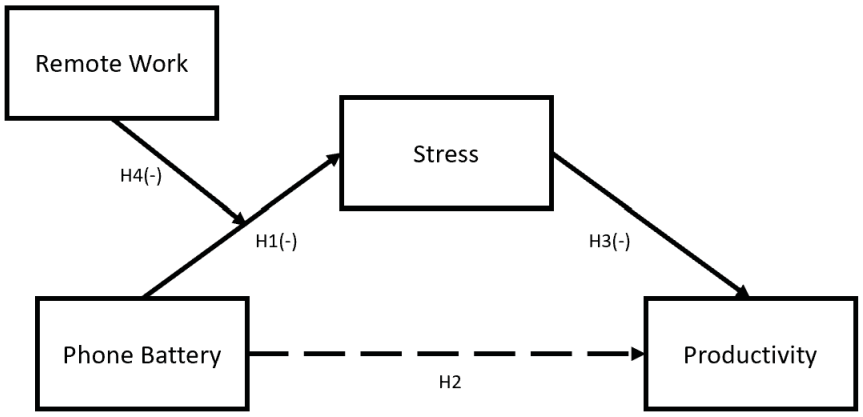


Figure 1: Conceptual Model

3. Data

3.1 The Tesserae Project

We used data from The Tesserae Project to explore this phenomenon. The Tesserae Project was a large-scale field experiment with 757 information workers across the US that intended to show how a suite of sensing devices could be used to measure workplace performance, psychological traits, and physical characteristics over a year (Mattingly et al. 2019). The dataset consisted of a wide variety of data that provided the framework for our consideration.

We received approval from the Project Lead Dr. Aaron Striegel and the entirety of the primary research team and completed a Data Usage Agreement to receive access to this data. The original project was approved by the University of Notre Dame IRB.

3.1.1 Mobile Sensing

Wearable: Each participant was given a Garmin Vivosmart 3, a fitness band-based smartwatch. The device was equipped with photoplethysmography (PPG) sensors for estimating heart rate and an accelerometer for tracking motion, step count, and physical activity. The

accuracy of the device's heart rate measurements has been corroborated through comparisons with electrocardiogram (ECG) readings and research-grade PPG sensors, showing a high correlation.

For heart-rate variability (HRV) assessment, the wearable provided both Root Mean Square of Successive Differences (RMSSD) and the Standard Deviation of the Average Normal-to-Normal (SDANN) intervals, calculated within specific time windows to enhance its precision. Using these measures, Garmin also provides a proprietary stress metric facilitated through Garmin's Health API that provides a fine-grained representation of fluctuating stress levels throughout the day.

The wearable also provided comprehensive tracking of physical activity. This included the duration of different activities, intensity levels, calories burned, step count, and distance traveled. The device provided both inferred summaries of user activities and allowed for manual entry, offering a detailed assessment of the participants' physical activity patterns. Finally, the wearable provided data on sleep duration, bedtimes, and wake times. The accuracy of sleep measurement was enhanced by adjusting the wearable-derived sleep data based on phone activity, addressing potential discrepancies in bed and wake time.

Beacons: Two types of Gimbal Bluetooth beacons were employed for data collection: Series 10 and Series 21. Series 10 beacons, characterized by their larger size and extended battery life of up to 18 months, were designated as static beacons. Participants placed these beacons at two fixed locations—their homes and workplaces. This placement ensured continuous monitoring of participant presence in these primary environments. Conversely, the Series 21 beacons, smaller in size and resembling key fobs, were designed for portability. Participants

could easily carry these in personal belongings such as wallets, purses, or laptop bags, allowing for mobility tracking.

The primary function of these beacons was to track participant location and activity though “beacon sightings.” These sightings, captured by the Gimbal API in conjunction with the participants’ smartphones, provided rich data sets that included timestamped information on beacon ID, signal strength, and ambient temperature. The static workplace beacon also allowed for the inference of break periods. The system categorized break sessions into intervals of 5, 15, and 30 minutes. Further, mobility location utilized the interplay between beacon sightings and smartphone location. The operational frequency of these beacons (2.4-GHz band) ensured a robust performance in diverse environments. This capability was essential for ensuring data accuracy in both home and work settings, which often present a range of physical and electronic obstacles. The effective range of up to 100 meters in unobstructed environments further enhanced the reliability of data collection.

Phone Agent: Participants installed a researcher-created app on their smartphones which tracked data usage, charging state and battery level, location, and screen locks/unlocks. The phone agent allowed for the calculation of distinct locations visited, distance between those locations, and the identification of significant locations for the participant (utilizing a DVSCAN clustering algorithm). It also synchronized with the wearable smartwatch to aide in the capturing of physical activity, ambient light levels, and other metrics. Further, the app interacted with the location beacons to log proximity data and map participant locations in relation to their homes or offices.

Miscellaneous Data: To provide further richness, various other data streams have been incorporated into the larger project. A subset of participants allowed the researchers to access

historical social media data from their Facebook and LinkedIn accounts. Additionally, daily weather data was collected from the World Weather Online Developer API using home zip codes. This data includes sunrise/sunset times, temperature, humidity, cloud cover, wind speed, visibility, and pressure.

3.1.2 Surveys

Ground Truth Measures: Participants were asked to complete a range of psychometric tools and self-reporting questionnaires, targeting an array of psychological, behavioral, and demographic variables. The comprehensive initial ground truth battery (IGTB) (Mattingly et al. 2019), took participants approximately 45–60 minutes to complete and included assessments across several domains. These include job performance, cognitive abilities, personality traits, mood, anxiety, health measures, and lifestyle factors such as exercise, sleep, and stress.

Demographics and Trait Measures: Participants reported demographic variables including age, sex, organizational role, nationality, education and income level. It additionally recorded trait measures including affect balance (Positive and Negative Affect Schedule [PANAS-X]) (Watson and Clark 1994), personality (Big Five Inventory [BFI-2]) (Soto and John 2017), sleep quality (Pittsburg Sleep Quality Index [PSQI]) (Buysse et al. 1989), sleep chronotypes (Morning-Eveningness Questionnaire [MEQ]) (Horne and Ostberg 1976), fluid and crystallized intelligence (Shipley 2) (Shipley et al. 2009), and anxiety (State Trait Anxiety Inventory (STAI-Trait Scale) (Spielberger 1983).

Work-Related Behaviors: Workplace performance was assessed across four different dimensions. Individual task performance (ITP) (Griffin et al. 2007) and in-role behavior (IRB) (Williams and Anderson 1991) reflect task performance and categorize duties or actions that are

formally recognized and rewarded by management. The former refers to the actions and behaviors that contribute to the production of a good or service where the latter reflects behavior that is required by an employee to accomplish their workplace duties.

Alternatively, organizational citizenship behavior (OCB) (Fox et al. 2012) and counterproductive work behavior (CWB) (Bennett and Robinson 2000) assess behaviors that promote the effectiveness of organizations. OCB includes actions that may not typically be rewarded or punished but still enhance workplace welfare (e.g., aiding a peer, volunteering in organization activities). Conversely, CWB reflects actions and behaviors that jeopardize the organization or colleagues (e.g., insulting a peer, stealing from the workplace). All four dimensions of workplace performance were captured in the IGTB.

Daily Surveys: To understand daily fluctuations in workplace performance, health, and well-being a series of short daily surveys were also distributed to participants. Participants received text messages that directed them to the surveys. They were intended to be completed within three minutes and were designed to assess various aspects of the participants' current state, including stress and anxiety.

Stress levels were gauged using a single question, rated on a 5-point Likert-type scale, asking participants to report the level of stress they were experiencing at that moment. The potential responses ranged from 'no stress at all' to 'a great deal of stress.' They were also asked an iterative cycle of questions (each daily survey asked the stress question and one of three other sets of questions) ranging from momentary assessments of work performance to sleep quality and health related outcomes. Surveys were sent at either 8AM, 12PM, or 4PM and participants were given 4 hours to respond.

3.1.3 Participants and Protocol

The study successfully recruited 757 participants, primarily from cognitively demanding professions including workers from consulting, engineering, information systems, and finance. Participants were drawn from across the United States. Recruitment was effectively conducted through multiple channels, including workplace emails, messaging boards, and newspaper advertisements. The inclusion of various university and corporate partners further enhanced the size of the participant base. The participant base was represented across genders, occupations, income levels, education, and job positions.

Participants were enrolled in the study between January and July of 2018, with data collection spanning just over a one-year period from January 2018 to March 2019. The participant pool was structured into four cohorts from four different institutions: a national technology services firm, a large US technology and engineering firm, a small US software firm, and a medium-sized US university. A fifth cohort consisted of miscellaneous interested applications recruited from unaffiliated channels (e.g., friends of recruited participants) The timeline of data collection and compensation is outlined in Figure 2 below.

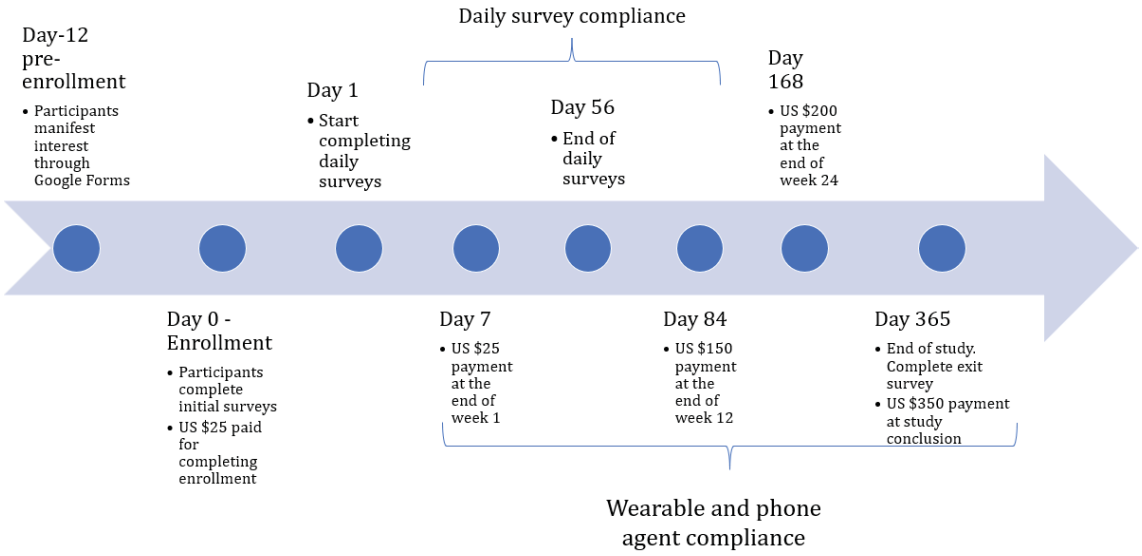


Figure 2: Data Collection Timeline

Each participant was equipped with the Garmin Vivosmart 3 devices and battery-powered Bluetooth location beacons described above. The longevity of these devices (e.g., 5-day battery life for the Garmin and 18-month functionality for the beacons) minimized participant intervention. The use of wearable technology and Bluetooth beacons provided a comprehensive method of data collection. Participants were instructed to wear the Garmin device continuously, barring charging and shower times, encouraging near-continuous data capture. The Bluetooth beacons were placed in the participant's home and at the participant's work desk. They were also instructed to always keep the key-fob beacons on them.

Given the sensitive nature of the data collected, the Tesserae Project placed a significant emphasis on participant privacy and data security. This was addressed through multiple avenues. First, participants provided written informed consent, in which high-level overviews of each sensing stream were supplemented with detailed technical specifications regarding the data collection and security measures. Further, the use of Open Authorization (OAuth) for the Garmin data collection ensured that personal details like usernames and passwords were not collected.

Data was anonymized and encrypted at various stages, form local storage on devices to transmission and eventual storage on secure servers. Randomized identifiers and HTTPS/SSH protocols were used for enhanced security. Finally, dividing front-facing and backend servers mitigated the risk associated with potential data breaches.

Figure 2 below provides a visual representation of the data that we had access to, and the process used to analyze that data:

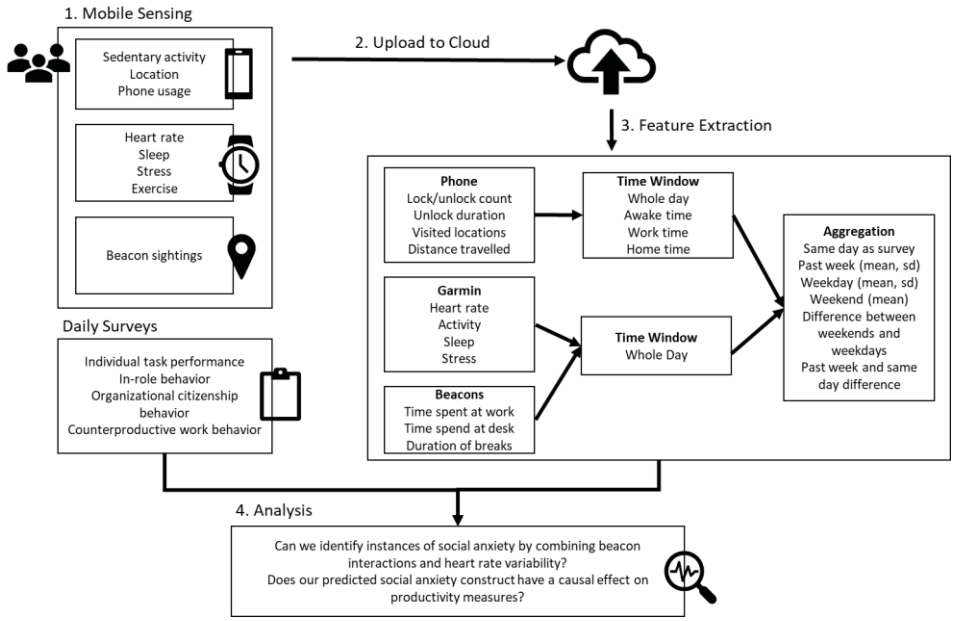


Figure 3: The Tesserae Project Data

3.2 Pre-Processing & Subset Selection

We restricted data collection to March 2018 through May 2018, providing three months of granular data. The first two months of the Tesserae Project (January through February) involved multiple batch enrollments and necessary adjustment periods to ensure the sensing streams were operating properly and participants were able to comply with the requirements of the study. Starting in June 2018, overall compliance began to drop. Therefore, the highest quality data with the lowest number of missing observations was collected over those three months.

Additionally, we removed participants who dropped out of the study during the designated window, who had a compliance measure two standard deviations below the mean, who had less than 50% compliance for daily survey completion, or who did not complete the IGTB. Our resulting dataset had 159 unique participants. Summary statistics for these participants is shown in Table 1.

Given the multimodal nature of the sensing streams, time increments differed across measures. The lowest time increment that would encompass all variations of data collection was 15-minutes. Therefore, data was grouped into 15-minute increments for each participant and the resulting features reflected the mean of all data collections that occurred within that window for that participant. For example, battery level could have been collected 20 times between 8 AM and 8:15 AM, but the resulting measure was the average battery level within that time frame. This resulted in 151,279 total observations.

Demographics	**Total Sample (n = 159)**
Age (years)	
Values, mean (SD)	34.18 (8.76)
Values, range	20–60
Sex, n (%)	
Male	124 (78%)
Female	35 (22%)
Income (US $), n (%)	
< 49,999	14 (9%)
50,000–74,999	32 (20%)
75,000–99,999	40 (25%)
100,000–150,000	57 (35%)
150,000+	18 (10%)
Education, n (%)	
No college degree	15 (9%)
College degree	104 (65%)
Graduate degree	40 (25%)
Supervisor role, n (%)	
Non–Supervisor	32 (20%)
Supervisor	127 (80%)
Personality (BIF–2)	

Neuroticism, mean (SD)	2.34 (0.75)
Conscientiousness, mean (SD)	3.80 (0.68)
Extraversion, mean (SD)	3.21 (0.68)
Agreeableness, mean (SD)	3.88 (0.55)
Openness, mean (SD)	3.82 (0.63)
Family Status	
Total dependents, mean (SD)	1.17 (1.37)
Dependents Under 1, mean (SD)	0.74 (0.54)
Dependents 2–5, mean (SD)	1.11 (0.62)
Dependents 5–12, mean (SD)	1.31 (0.79)
Dependents 13+, mean (SD)	1.26 (0.86)

Table 1: Participant Summary Statistics

4. Analysis

4.1 Directed Acyclic Graph (DAG)

Causal diagrams, in the form of directed acyclic graphs (DAGs), allow for the proper encoding of relatively precise theoretical claims around the effect of one variable on another. DAGs help in avoiding the introduction of bias when determining the proper set of control or mediating variables and in identifying sources of confoundedness (Tafti and Shmueli 2020). We utilized this practice to serve as a conceptual aid to find, among the full set of available features, which subset of those features are sufficient for identification, and which are optimal for inclusion in the model.

We started by setting the treatment node, phone-battery level (PBL), and outcome node, stress level (SL). The next step was to add nodes for any measured variables available in the data. However, given the extensive number of features available to us, we first identified only those features that were likely to influence PBL or SL. These included individual traits (IT) such as trait anxiety, personality type, trait affect, and emotional intelligence. We also included external factors such as personal life events. The daily surveys included an open-ended question that allowed participants to report any adverse or unexpected life events to the researchers.

Next, we identified any unobserved features that may lie on the spurious path from any of the nodes to the outcome. Most notably, this included phone usage habits (PUH). Although data was collected on the number of phone locks and unlocks, potentially giving us a measured way to control PUH, changes in permission requirements from Android and iOS operating systems made this collection infrequent and unpredictable. Given the lack of confidence in data quality, we assumed that this was unobserved. Additionally, PUH may also consider whether a mobile phone is being used for work or personal purposes, which is again unobserved in our data. Finally, we also included any temporal effects as a node to account for daily fluctuations in PBL and SL that may correlate over time within individuals.

The final step in constructing the DAG was to add arrows to and from each node when theory or logic would have lead us to believe that a direct causal effect may exist. A direct line is shown from PBL to SL reflecting our H1 described above. For elements included in IT, we looked to prior literature to find evidence for any causal effect. Numerous studies have highlighted the correlation between anxiety and smartphone use (Gao et al. 2016; Konok et al. 2016; Lepp et al. 2014) and anxiety and stress (Bystritsky and Kronemyer 2014; Fedoce et al. 2018). Similar findings exist for personality traits (Marengo et al. 2020; Stachl et al. 2020) and emotional intelligence (Arrivillaga et al. 2020; van Deursen et al. 2015). We denoted the arrows from IT to SL and PBL as partially observed. Although we have several important traits measured in our data, we did not assume that this list is exhaustive.

Given the description provided in *Section 2.1* on the dual role of mobile phones as both a stressor and a comforter (Melumad and Pham 2020), it is likely that PUH introduces significant complexity to our identification. Utilizing mobile phones for leisure activities can offer psychological detachment from work or other stressors (Lanaj et al. 2014), thereby reducing PBL

while simultaneously reducing SL. Alternatively, utilizing mobile phones for sustained work duties would similarly reduce PBL while also increasing SL. To properly identify our causal effect of interest, this backdoor had to be accounted for. We also included a direct effect of time on PBL and SL to account for any temporal effects that might have biased our results. Finally, we included direct effects of EF (personal life events) on both stress and battery life. Take, for instance, an individual who is awaiting consequential medical results. They are likely to increase their frequency of checking their phone (thereby decreasing PBL) while also increasing SL from their health-related worries.

Our resulting DAG is shown in Figure 3 below. The following direct and indirect paths exist between PBL and SL:

1. PBL → SL (the causal effect of battery level on stress)
2. PBL ← IT → SL (backdoor path from individual traits)
3. PBL ← PUH → SL (backdoor path from phone usage habits)
4. PBL ← EF → SL (backdoor path from external factors)
5. PBL ← T → SL (backdoor path from temporal effects)

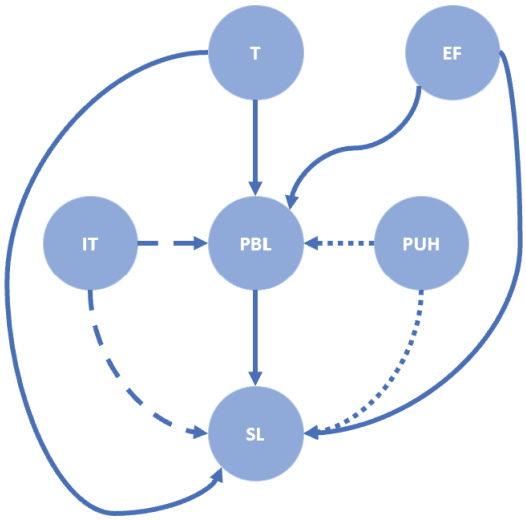


Figure 4: Directed Acyclic Graph for the Causal Effect of Phone-Battery Level on Stress

4.1 Estimation Approach

Our first objective is to identify the causal effect of phone-battery level on stress. To test this empirical relationship, we start with the following individual-panel specification:

$$SL_{it} = \beta_0 + \beta_1 PBL_{it} + \beta_2 X_{it} + \varepsilon_{it}$$

where subscripts refer to individual i in 15-minute time increment t. SL reflects the participant's level of stress and PBL reflects the participant's phone-battery level. X denotes a matrix of other covariates that influence stress independently of battery level. These include various measures of physical activity in the given time-increment as that is likely correlated with the HRV-based stress score. Finally, ε captures unobserved determinants that vary across individuals and time increments.

However, in the preceding section we identified several backdoor paths that must be closed to identify the true causal effect of PBL on SL. Given that IT is time-invariant, the use of participant fixed effects (equivalent to demeaning PBL and SL) in our model would account for any effects of the above-mentioned traits by eliminating any time-invariant, individual-specific

heterogeneity. Alternatively, EF is time-varying which, given that the inclusion of fixed effects would simply demean a time-variant variable, moving its effect into the composite error term that is correlated with PBL. However, because EF is also observed, we simply control for it as a time-varying feature in our model. We can then update our specification such that:

$$SL_{it} = \beta_0 + \beta_1 PBL_{it} + \beta_2 EF_{it} + \beta_3 X_{it} + \alpha_i + \varepsilon_{it}$$

where EF reflects reported personal life events for the participant[21] and α captures unobserved time-invariant individual fixed effects (IT).

Two remaining issues must be addressed in our estimation strategy: temporal effects and endogeneity related to phone usage habits. For example, it is possible reverse causation exists in which people may experience an increase in SL caused by a decrease in PBL *or* increased SL that affects PUH, which subsequently decreases PBL. Given that PUH is unobserved, addressing the endogeneity that stems from its effect is more complex. To address these two issues, we use dynamic panel estimation techniques. Utilizing lagged outcome variables in the model specification could account for temporal effects but not the issue of endogeneity. To do this, a common approach is to use the method first introduced by Arellano and Bond (1991) who suggested a first-differenced GMM (Diff-GMM). This method deals with the problem of endogeneity by first differencing the data and then utilizing suitably lagged values of the independent and dependent variables as instruments. However, under cases of heteroskedasticity, this one-step GMM estimator is no longer efficient. Therefore, a two-step Diff-GMM that utilizes the optimal weighting matrix resulting in a significant reduction in standard error bias, is preferred.

[21] If a participant reported a personal life event on day T, that would be indicated for any time period $t \in T$.

Later work extended this methodology to account for the weak instrument problem with the development of the system GMM estimator (Sys-GMM) which has been shown to increase efficiency considerably (Blundell and Bond 1998). Although this method is more efficient, it does not entirely rectify the weak instrument problem. Therefore, we also estimate the model utilizing the number of location readings (LR) as an instrument. Going back to our DAG, utilizing LR as an instrument would imply that there is a mediated pathway from LR to SL via PBL. Namely, as the number of location readings varies, stress too should vary, but only because battery level is varying. In this specification, LR is independent of our unobserved variable PUH, satisfying the exclusion restriction. PBL becomes a collider along the path LR → PBL ← PUH → SL, allowing for the identification of our causal effect of interest.

To justify our claim that LR is independent of PUH and only causes variations in SL through its effect on PBL, we examine the process by which location readings were collected during the Tesserae Project. Whereas location data was collected via sightings of the Bluetooth beacons at the participant's home and work, a secondary stream of location data was also collected through the phone agent accessing the device's GPS. The mechanics of mobile operating systems require several complex workarounds from application developers which are mostly beyond the scope of this work. However, one relevant workaround was designed to ensure that the application could operate as a background process on the participant's mobile phone, even if the phone was not being used for extended periods of time. This workaround led to randomly timed readings of GPS coordinates through the phone agent. As the readings were not dependent on the participant's behavior and were strictly a function of random background processes, we can confidently assume that the count of said readings is uncorrelated with stress. Equally important, because the collection of location readings required an output of energy from

the participant's device, doing so *decreased phone-battery level*. This makes LR a suitable instrument for handling the endogeneity in our model.

Estimating the above models allowing for the optimal weighting matrix creates several empirical challenges. First, the model likely becomes strongly overidentified when the inclusion of a large number of instruments overfits the instrumented variables (Roodman 2009). Further, instrument proliferation can lead to significant under rejection of overidentification tests, thereby incorrectly signaling that the model is correctly specified. Additionally, given the large number of time periods in our sample, it is computationally infeasible to estimate these models unconstrained. Therefore, we utilize curtailing as a method of restricting the number of lags, in line with Roodman (2009).

4.1.1 Results

To facilitate comparison with common techniques used to identify causal relationships, we first estimate the above equation using random and fixed effects. These results are reported in Table 1. Estimates in Column (1) are outputs of a simple panel random effects model with clustered standard errors. These results reflect a negative and statistically significant correlation between phone-battery level and stress. This is consistent with our theory that the two features would, at the least, vary together. The results in Column (2) attempt to account for some temporal effects by including time fixed effects in the estimate. These results show a similarly negative relationship between phone-battery level and stress but to a lesser severity and a slightly lower significance. This result implies a high likelihood of time effects in our underlying data-generation process in which, for instance, as the day progresses, phone-battery level naturally decreases and stress levels may change, as well.

The results in Column (3) add the individual fixed effects in conjunction with time fixed effects. Interestingly, the estimate for the effect of phone-battery level on stress does not change in direction, severity, or significance from Column (2). However, examining the R^2 shows a large jump in the amount of information captured by the model. This result reveals that although individual time-invariant characteristics are important for explaining stress, they do not influence the relationship between phone-battery level and stress. The results in Column (4) involve estimates of the same model as Column (3), with the addition of four covariates: physical activity (in seconds), high intensity physical activity (in seconds), step count, and the presence of any reported personal life events. Again, our estimate for the effect of battery level remains mostly unchanged (with the exception of a slight decrease in severity). We find positive and statistically significant effects of both physical activity covariates. This is aligned with our understanding of the HRV-based estimate of stress utilized by Garmin. Given that HRV can rise because of either stress or physical activity, it is important to control for these features. Interestingly, step count has a negative relationship with stress. This result likely occurred because steps do not necessarily increase HRV, especially if they are merely registered from casual walking, but likely decrease stress.

	RE (1)	RE + Time (2)	RE + Time/Unit (3)	RE + ALL (4)
Battery Level	-0.046***	-0.025**	-0.025***	-0.022***
	(0.009)	(0.008)	(0.008)	(0.008)
Active				0.049***
				(0.002)
Highly Active				0.050***
				(0.004)
Step Count				-0.024***
				(0.002)
Life Event				1.23
				(1.28)

N	151,279	151,279	151,279	151,279
R^2	0.0043	0.0220	0.2683	0.3063

Table 2: Phone-battery level and Stress - RE, IV, and Fixed Effects Estimates

Table 3 reports estimates from our attempts to control endogeneity and close all backdoor paths highlighted in our DAG. Columns (1)–(3) report our GMM estimates. Column (1) reports the one-step Diff-GMM estimates. These results show a highly significant effect of the lagged dependent variable, highlighting the role of temporal effects in our data. The effect of battery level is large and negative but statistically insignificant. Covariate effects are consistent with theory. However, this estimator is not robust to serial correlation. Higher-order effects of serial correlation appear to be present given the low AR(2) p-value. This implies that coefficients in Column (1) are not consistent. Column (2) provides estimates from the two-step Diff-GMM. The effect sizes and directions are consistent with Column (1) with only slight variations. Both Diff-GMM models pass the overidentification tests.

The instruments used for the first two estimators are likely weak. Therefore, Column (3) estimates the Sys-GMM estimator. These results show a much lower effect of battery level on stress (compared to Diff-GMM) but the effect is now statistically significant. While results of the AR tests imply that serial correlation is a less significant issue than it was for the Diff-GMM models, the overidentification tests fail for this estimator.

Finally, Column (4) reports the most robust of our analyses, utilizing the number of location readings as an instrument for battery level. The test for endogenous regressors was run and reported a Chi-Squared p-value of 0.0056 implying the presence of endogeneity in our model and confirming the need for instrumentation. Testing for weak instruments reports a Wald *F* statistic of 57.19, significantly larger than the standard threshold of 10 for a single instrument

regression. Further, testing for under identification reports a Chi-Squared p-value of 0.000. Overall, these results confirm that we have a strong instrument. This estimator addresses each of the backdoors highlighted by our DAG, utilizing fixed effects, the inclusion of a lagged dependent variable in the model, controlling for observed time-variant confounders, and addressing unobserved confounders through instrumentation. These results show a strong positive effect of the lagged stress score and a modest but statistically significant negative effect of battery life. This implies that decreasing phone-battery level has a causal effect on stress levels, providing evidence in support of H1 and highlighting the effects of nomophobia.

	Diff-GMM-1 (1)	Diff-GMM-2 (2)	Sys-GMM (3)	IV (5)
Stress Lag	0.501***	0.502***	0.516***	0.710***
	(0.015)	(0.015)	(0.019)	(0.007)
Battery Level	-0.347	-0.338	-0.120*	-0.061***
	(0.220)	(0.215)	(0.072)	(0.018)
Active	0.028***	0.018***	0.020***	0.019***
	(0.002)	(0.002)	(0.002)	(0.001)
Highly Active	0.010***	0.010***	0.018***	0.018***
	(0.004)	(0.004)	(0.005)	(0.002)
Step Count	-0.009***	-0.009***	-0.011***	-0.007
	(0.001)	(0.002)	(0.002)	(0.001)
Life Event	-	-	29.31*	0.100
	-	-	(0.072)	(0.423)
n	109,184	109,184	109,184	151,279
AR(1) *p*	0.0000	0.0000	0.0000	-
AR(2) *p*	0.0598	0.0594	0.1034	-
Hansen J *p*	0.7308	0.8453	0.0479	-
Sargan *p*	0.7928	0.8451	0.0196	-

Table 3: Phone-battery level and Stress - GMM Estimates

4.2 Objective 2 – Estimation Approach

Our second objective is to explore the effect of phone-battery level on productivity, utilizing stress levels as a mediator. To do this, we utilize structural equation modeling as a way

to parse out the direct and indirect effects of phone-battery level. We use the following two regression equations:

$$M = \beta_{0M} + \beta_1 T + \beta_2 \boldsymbol{X} + \varepsilon_M$$

$$Y = \beta_{0Y} + \beta_3 M + \beta_4 T + \beta_5 \boldsymbol{X} + \varepsilon_Y$$

where M reflects our mediator (stress), T reflects battery level and $\boldsymbol{X}$ is a vector of covariates including our physical activity indicators and time fixed effects. We estimate the model for four different variations of Y: individual task performance (ITP), in-role behavior (IRB), organizational citizenship behavior (OCB), and counterproductive work behavior (CWB). β_1 is the path coefficient for the effect of battery level on stress, β_3 is the path coefficient for the effect of stress on job performance, and β_4 is the path coefficient for the effect of battery level on job performance (reflecting the direct effect). Therefore, we can calculate indirect effects of phone-battery level on performance, mediated through stress, as $\beta_3 \times \beta_1$ and total effects as $\beta_4 + \beta_3 \times \beta_1$.

4.2.1 Results

Results of our analysis are shown in Table 4 below. The direct effect of phone-battery level on productivity. We report the indirect, direct, and total effect of phone-battery level on each of the four performance measures. Indirect effects of phone-battery level mediated through stress show small but statistically significant decreases to task-related performance measures (ITP, IRB). This implies that as battery level decreases, it creates an increase in stress that subsequently lowers task performance. This provides evidence in support of H3 that battery level has a negative indirect on productivity. CWB similarly has a negative and significant indirect effect, while indirect effects for OCB are insignificant.

Type	Effect	Estimate	SE	95% C.I. Lower	95% C.I. Upper	z	p
Indirect	$\beta_3\beta_1$						
ITP		-0.0001***	0.00002	-0.00014	-0.00006	-5.08	0.000
IRB		-0.0003**	0.000	-0.00056	-0.00001	-2.06	0.039
OCB		0.00002	0.00003	-0.00004	0.00007	0.59	0.557
CWB		-0.0004***	0.00005	-0.00045	-0.00026	-7.60	0.000
Direct	β_4						
ITP		0.0022***	0.0003	0.00150	0.00290	6.07	0.000
IRB		0.0195***	0.0020	0.01523	0.02381	8.92	0.000
OCB		0.0029***	0.0005	0.00200	0.00379	6.25	0.000
CWB		-0.0043***	0.0004	-0.0052	-0.00346	-9.88	0.000
Total	$\beta_4 + \beta_3\beta_1$						
ITP		0.0019***	0.0003	0.00139	0.00247	7.02	0.000
IRB		0.0192***	0.0020	0.01495	0.02351	8.80	0.000
OCB		0.0029***	0.0005	0.00230	0.00381	6.29	0.000
CWB		-0.0047***	0.0004	-0.00553	-0.00381	-10.67	0.000

Table 4: Mediation Analysis – Phone-battery level & Performance

Direct effects of the four performance measures are all moderately sized and statistically significant. This implies that the direct effect of decreasing phone-battery level is performance *increasing*. This supports H2a. It could be the case that decreasing phone-battery level leads to employees checking their phones less often to preserve battery and therefore decreasing some of the attention hurdles that negatively influence productivity.

4.3 Objective 3 – Estimation Approach

In the previous section, we saw that nomophobia (operationalized as phone-battery level mediated stress) had a negative effect on workplace performance. To further uncover heterogenous effects around this phenomenon, we introduce a moderator to our mediation analysis that reflects the remote work status of the participants. We use the following equations to model the indirect moderated and unmoderated effects:

$$M = \beta_{0M} + \beta_1 T + \beta_2 W + \beta_3 (T \times W) + \beta_4 \boldsymbol{X} + \varepsilon_M$$

$$Y = \beta_{0Y} + \beta_5 M + \beta_6 T + \beta_6 W + \beta_7 (T \times W) + \beta_8 \boldsymbol{X} + \varepsilon_Y$$

where W is our moderating variable, remote work status. Our conditional indirect effect of phone-battery level on performance can then be quantified as $(\beta_1 + \beta_3) \times \beta_5$. We focus our outcome strictly on IRB given that our initial mediation results were most significant for this measure.

4.3.2 Results

Results of our moderated-mediation analysis are shown in Table 5 below. Our indirect effects show that for in-person work, phone-battery level has a negative indirect effect on performance mediated through stress. Interestingly, this effect goes away when conditioned on working from home. This implies that the negative implications of nomophobia on productivity may be attenuated by remote work, providing support for H4. Our direct effects show similar results to our mediation analysis, that phone-battery level decreasing may actually increase productivity.

				95% C.I.			
Type	Effect	Estimate	SE	Lower	Upper	z	p
Indirect	$(\beta_1 + \beta_3) \times \beta_5$						
Remote		-0.0001	0.0001	-0.00027	0.00016	-0.49	0.622
In-Person		-0.0004***	0.0002	-0.0007	-0.0001	-2.64	0.008
Direct	$(\beta_6 + \beta_7)$						
Remote		0.0020	0.0059	-0.009	0.0135	0.34	0.735
In-Person		0.0225***	0.0024	0.0179	0.0271	9.54	0.000
Total	$[(\beta_1 + \beta_3) \times \beta_5] + [(\beta_6 + \beta_7)]$						
Remote		-0.0068	0.0136	-0.0335	0.0198	-0.50	0.616
In-Person		-0.0546***	0.0055	-0.0653	-0.0439	-10.00	0.000

Table 5: Moderated-mediation Analysis – Remote Work

4.3 Robustness

Although wearables provide us with a large amount of valuable data with fine granularity, the repeated measures pose additional challenges. To address these shortcomings and provide further robustness for our measured relationship between changes in phone-battery level and stress, we utilize a human-in-the-loop segmented mixed-effects modeling method proposed by Srinivasan et al. (2023). This method utilizes a human-in-the-loop approach to identify change points in a segmented model (a combination of an algorithmic search process that conducts fine-tuning through human inputs). The model is fit with the following equation:

$$y_{ij} = \beta_0 + \gamma_{0j} + \sum_{s \in S} \beta_r^{(s)} x_{rij} \times I(x_{rij} \in s) + \sum_{k=1, k \neq r}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij} + \epsilon_{ij}$$

where S a segment set based on the change points identified for x_r. We can measure the significance of the effect on the input variable, in our case change in phone-battery level, at each segment, s, by estimating the fixed-effects coefficient $\beta_r^{(s)}$.

We start by fitting a generalized additive mixed model (GAMM) to allow for a visual inspection of the smooth function with the goal of identifying initial estimates of change points (P) in the data. We include additional covariates for physical activity as we did in *Section 4.1*. Following the visual estimation of the change points, we use Brent's method (a root-finding algorithm) to maximize the mixed-effects model's Akaike information criteria and identify precise change points. This process utilizes human inputs to improve the process of fitting segmented models, allowing us to estimate higher-order associations between changes in battery level and stress.

4.3.1 Results

The estimation of our GAMM model can be seen in Figure 3 below. The results show two distinct extrema. A maxima exists for $Battery$ around 20% followed by a decrease in the smooth function, reaching a stopping point around $Battery = 63\%$. The function continues to decrease towards another minima around 80% before increasing to a maxima at 89%. Given these results, we specify P = 4 in our estimation of Brent's method with estimated break points at $Battery = 20, Battery = 63, Battery = 80$ and $Battery = 89$.

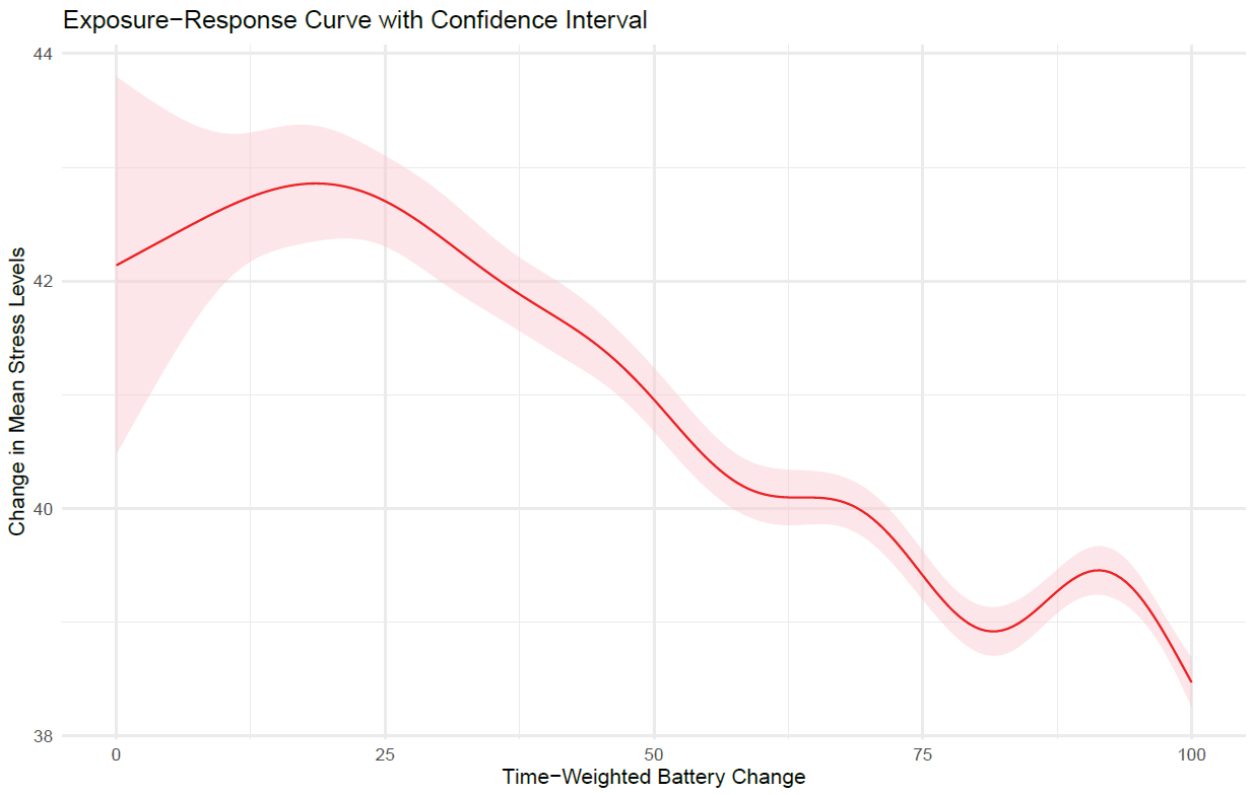


Figure 3: Smooth Function Plots Generated Using GAMM for Stress-Battery

Along with generating our mixed-effects segmented model, we estimate a standard linear and a curvilinear model to serve as benchmarks. Comparing our results to the linear model highlight the tradeoff between simplicity and better fit from identifying higher-order associations

whereas the curvilinear improves fit but has lower interpretability of higher-order coefficients. The fixed effects of segmented inputs are shown in Table 4 below. In the linear and curvilinear models, we see modest effects of battery level on stress, implying that larger drops in battery level result in higher increases to stress. Estimates from our segmented model utilize breakpoints at 24, 58, 76, and 91 (identified via the combination of human input and Brent's method). Our results are shown in Table 4 below. First, we find that the linear and curvilinear models overestimate the effect of battery level on stress. By segmenting our model and controlling for nuance around natural breakpoints in the data, we are able to observe several interesting results. First, we see initial increases in stress from battery levels dropping from 100% to ~91%. Then after crossing the battery level threshold at 76%, increases in stress accelerate. The largest jump in stress occurs between 58% and 24%, at which point increases to stress plateau. We were able to isolate this effect and provide further specification to our results by estimating the segmented model.

Outcome	**Input**	**Coefficient**		
		Segmented	**Linear**	**Curvilinear**
Stress	$Battery$		-0.051 (0.008)***	-0.043 (0.014)***
	$Battery^2$			-0.001 (0.00)***
	$Battery < 24\%$	-0.008 (0.035)		
	$24\% < Battery \leq 58\%$	-0.028 (0.035)***		
	$58\% \leq Battery \leq 76.4\%$	-0.032 (0.006)***		
	$76.4\% \leq Battery \leq 90.6\%$	-0.038 (0.005)***		
	$90.6\% \leq Battery$	-0.036 (0.005)***		

Table 4: Fixed Effects of Segmented, Linear, and Curvilinear Models

5. **Discussion & Conclusion**

Our results provided a number of important insights. First, motivated by the development of the nomophobia construct, we examined the effect of phone battery level on employee stress levels throughout the work day. Descriptive and correlational analysis highlights a moderate

negative relationship between the two features, implying that as battery level decreases, stress levels increase. To control for endogeneity and identify the causal effect of phone battery on stress, we utilize a number of econometric techniques. We utilize three different GMM estimators for dynamic panel data which show much larger negative effects of phone battery than originally thought. However, due to concerns around overidentification and weak instruments, we find that these estimates are not reliable. To confidently isolate our effect, we use the number of location sightings as an instrumental variable for battery level. These results show a 0.06 increase in stress levels for every one percentage decrease in battery levels. Stress scores across our sample have a mean of 38 and a standard deviation of 16. This implies that, on average, for every battery decrease of 10% increases stress by around 2%.

Our robustness checks further explore this effect by allowing for segmented analyses at different thresholds of battery level change. We find break points in the data that allow us to identify thresholds in the effect of battery level on stress. We find that once battery levels drop below ~58%, increases in stress accelerate. Further, we find that battery-related stress increases plateau once battery level drops below 24%.

Further we conduct mediation analysis to explore the role of nomophobia (operationalized as the indirect effect of phone battery through stress) on performance. First, we see a positive direct effect of phone battery on performance. This tells us that decreasing phone battery improves performance. It could be the case that drains on employee attention resulting from more conservative utilization of mobile phones is the driver of this effect. Second, we find that overall, phone battery-related stress does have a modest and significant indirect effect on performance that implies that as phone battery decreases, stress levels go up, which subsequently causes performance to diminish. Finally, we perform a moderated mediation estimation that

looks at the role of remote work status in moderating the relationship between phone battery and stress. We find that remote work attenuates all of the measured indirect effect that's observed with in-person work. This implies that the negative implications of nomophobia on productivity may not be as severe when working from home.

This work has a number of important implications for academia. At a high level, we contribute to the IS use literature that considers the negative implications of technology use and dependence. Our results add further nuance to this work, highlighting the role of nomophobia in influencing stress levels. Further, we contribute to the clinical literature on nomophobia, utilizing a large-scale longitudinal data set that allows us to measure direct evidence of nomophobia implications in a way that the majority of prior work has not been able to do. The use of the Tesserae Project data allows for a robust operationalization of our research questions, allowing for a more nuanced consideration of society's complex relationship with mobile phones. We also utilize interdisciplinary theoretical frameworks to motivate both our research questions and our methodological design.

Given the growing interest in studying the effects of remote work in light of the Covid-19 pandemic, our work contributes important insight to the field. We show the positive effects of remote work status on nomophobia-related performance decreases. Future work could further explore this relationship, uncovering the drivers behind this finding. It could be that remote workers have easier access to charging capabilities resulting in a comfort that if their battery was to get too low, they would be able to easily address it. It also could be that the way in which phones are used differs between work and home, leading to a change in the effects of nomophobia.

Further, we provide a number of novel methodological contributions to the literature. Specifically, we answer the call of Tafti and Shmueli (2020) to utilize a robust design framework for causal diagrams to better motivate IS research. We use their framework to identify important backdoor paths in our estimations and devise identification strategies that close those paths. We also utilize a novel human-in-the-loop mixed methods approach in our robustness checks that captures the benefit of human and computer collaboration in identifying higher-order relationships between features. This methodology, developed by Srinivasan et al. (2023), was designed specifically for the use of wearable data and therefore IS work that utilizes such data could benefit from the design immensely. Finally, we utilize a robust instrumental variable that allows us to isolate causal effects from multimodal sensor data, an incredibly complex task. Our IV, while unintentionally put in place by the researchers, could also inform the design of future field work that utilizes the randomness of backend technology to exogenously manipulate features of interest.

This work is not without limitations. First, as mentioned, multimodal sensing data is incredibly complex and isolating causal effects is not a straightforward task. While we are confident in our identification strategy, it could be that other confounding factors exist in the relationship between phone battery and stress. The use of secondary field data, while increasing external validity, decreases the control we have over the environment and the precision with which we can examine the results. Further, our participants did not complete a standardized nomophobia questionnaire, something that would have surely lent further insight to our work. While we cannot make definitive claims that behavior was specifically nomophobia, as defined in the literature, we do feel comfortable that our operationalization was robust enough to still lend important insights to the field.

Our results also have important implications for industry. First, we contextualize the study of nomophobia to the workplace because of the important balance between firm performance and employee well-being. Helping organizations to better understand their own employees and their relationship to mobile devices in the workplace will aide in making more informed managerial decisions, balancing the good of their employees and the good of the firm. Further, highlighting the benefits of remote work may help inform "return-to-work" strategies in a post-COVID world. Further, policymakers may benefit from this work, better understanding the stakes of organizational decision making in encouraging or discouraging the use of mobile phones in the workplace. While a complex matter, highlighting the implications on employee mental health can better support proposed regulation. Individuals can also benefit from a more robust understanding of their own relationship to their mobile phones, ensuring the proper attention is paid to their own well-being.

6. References

Aguilera-Manrique, G., Márquez-Hernández, V. V., Alcaraz-Córdoba, T., Granados-Gámez, G., Gutiérrez-Puertas, V., and Gutiérrez-Puertas, L. 2018. "The Relationship between Nomophobia and the Distraction Associated with Smartphone Use among Nursing Students in Their Clinical Practicum," *PLOS ONE* (13:8), p. e0202953.

Ahmed, S., Pokhrel, N., Roy, S., and Samuel, A. J. 2019. "Impact of Nomophobia: A Nondrug Addiction among Students of Physiotherapy Course Using an Online Cross-Sectional Survey," *Indian J Psychiatry* (61:1), pp. 77-80.

Aldhahir, A. M., Bintalib, H. M., Siraj, R. A., Alqahtani, J. S., Alqarni, O. A., Alqarni, A. A., Alghamdi, H. S., Alyami, M. M., Naser, A. Y., Fatani, A. I., and Alwafi, H. 2023. "Prevalence of Nomophobia and Its Impact on Academic Performance among Respiratory Therapy Students in Saudi Arabia," *Psychology Research and Behavior Management* (Volume 16), pp. 877-884.

Amichai-Hamburger, Y., and Vinitzky, G. 2010. "Social Network Use and Personality," *Computers in human behavior* (26:6), pp. 1289-1295.

Arellano, M., and Bond, S. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* (58:2), pp. 277-297.

Arrivillaga, C., Rey, L., and Extremera, N. 2020. "Adolescents' Problematic Internet and Smartphone Use Is Related to Suicide Ideation: Does Emotional Intelligence Make a Difference?," *Computers in Human Behavior* (110), p. 106375.

Ayyagari, R., Grover, V., and Purvis, R. 2011. "Technostress: Technological Antecedents and Implications," *MIS Quarterly* (35:4), pp. 831-858.
Bartwal, J., and Nath, B. 2020. "Evaluation of Nomophobia among Medical Students Using Smartphone in North India," *Medical Journal Armed Forces India* (76:4), pp. 451-455.
Bennett, R. J., and Robinson, S. L. 2000. "Development of a Measure of Workplace Deviance," *Journal of applied psychology* (85:3), p. 349.
Blundell, R., and Bond, S. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* (87:1), pp. 115-143.
Boswell, W. R., and Olson-Buchanan, J. B. 2007. "The Use of Communication Technologies after Hours: The Role of Work Attitudes and Work-Life Conflict," *Journal of management* (33:4), pp. 592-610.
Bragazzi, N., and Del Puente, G. 2014. "A Proposal for Including Nomophobia in the New Dsm-V," *Psychology Research and Behavior Management*), p. 155.
Bragazzi, N. L., Re, T. S., and Zerbetto, R. 2019. "The Relationship between Nomophobia and Maladaptive Coping Styles in a Sample of Italian Young Adults: Insights and Implications from a Cross-Sectional Study," *JMIR Mental Health* (6:4), p. e13154.
Brunborg, G. S., Mentzoni, R. A., Molde, H., Myrseth, H., Skouverøe, K. J. M., Bjorvatn, B., and Pallesen, S. 2011. "The Relationship between Media Use in the Bedroom, Sleep Habits and Symptoms of Insomnia," *Journal of sleep research* (20:4), pp. 569-575.
Butts, M. M., Becker, W. J., and Boswell, W. R. 2015. "Hot Buttons and Time Sinks: The Effects of Electronic Communication During Nonwork Time on Emotions and Work-Nonwork Conflict," *Academy of Management Journal* (58:3), pp. 763-788.
Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. 1989. "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research," *Psychiatry research* (28:2), pp. 193-213.
Bystritsky, A., and Kronemyer, D. 2014. "Stress and Anxiety: Counterpart Elements of the Stress/Anxiety Complex," *Psychiatric Clinics* (37:4), pp. 489-518.
Chang, L. 2015. "70 Percent of Drivers Use Their Smartphones While on the Road." from https://www.digitaltrends.com/mobile/70-percent-of-us-use-our-phones-while-driving/
Chen, A., and Karahanna, E. 2018. "Life Interrupted: The Effects of Technology-Mediated Work Interruptions on Work and Nonwork Outcomes," *Mis Quarterly* (42:4), pp. 1023-+.
Choudhary, S., and Mishra, K. 2023. "Understanding Knowledge Hiding in the Context of Virtual Workplaces," *VINE Journal of Information and Knowledge Management Systems* (53:3), pp. 566-589.
Condliffe, J. 2017. "Constant Phone Checkers Are Totally Strung Out." from https://www.technologyreview.com/s/60370/constant-phone-checkers-are-totally-string-out/
D'Arcy, J., Gupta, A., Tarafdar, M., and Turel, O. 2014. "Reflecting on the "Dark Side" of Information Technology Use," *Communications of the Association for Information Systems* (35:1), p. 5.
DailyMail. 2008. "Nomophobia Is the Fear of Being out of Mobile Phone Contact - and It's the Plague of Our 24/7 Age." from https://www.dailymail.co.uk/news/article-550610/Nomophobia-fear-mobile-phone-contact--plague-24-7-age.html
Dellarocas, C., Sutanto, J., Calin, M., and Palme, E. 2015. "Attention Allocation in Information-Rich Environments: The Case of News Aggregators," *https://doi.org/10.1287/mnsc.2015.2237*).

Derks, D., and Bakker, A. B. 2014. "Smartphone Use, Work–Home Interference, and Burnout: A Diary Study on the Role of Recovery," *Applied Psychology* (63:3), pp. 411-440.
Einav, L., Levin, J., Popov, I., and Sundaresan, N. 2014. "Growth, Adoption, and Use of Mobile E-Commerce," *American Economic Review* (104:5), pp. 489-494.
Elhai, J. D., Dvorak, R. D., Levine, J. C., and Hall, B. J. 2017. "Problematic Smartphone Use: A Conceptual Overview and Systematic Review of Relations with Anxiety and Depression Psychopathology," *Journal of Affective Disorders* (207), pp. 251-259.
Farooqui, I. A., Pore, P., and Gothankar, J. 2018. "Nomophobia: An Emerging Issue in Medical Institutions?," *Journal of Mental Health* (27:5), pp. 438-441.
Fedoce, A. d. G., Ferreira, F., Bota, R. G., Bonet-Costa, V., Sun, P. Y., and Davies, K. J. A. 2018. "The Role of Oxidative Stress in Anxiety Disorder: Cause or Consequence?," *Free Radical Research* (52:7), pp. 737-750.
Ferreira, D., Dey, A. K., and Kostakos, V. 2011. "Understanding Human-Smartphone Concerns: A Study of Battery Life," *Pervasive Computing: 9th International Conference, Pervasive 2011, San Francisco, USA, June 12-15, 2011. Proceedings 9*: Springer, pp. 19-33.
Fox, S., Spector, P. E., Goh, A., Bruursema, K., and Kessler, S. R. 2012. "The Deviant Citizen: Measuring Potential Positive Relations between Counterproductive Work Behaviour and Organizational Citizenship Behaviour," *Journal of Occupational and Organizational Psychology* (85:1), pp. 199-220.
Galanti, T., Guidetti, G., Mazzei, E., Zappalà, S., and Toscano, F. 2021. "Work from Home During the Covid-19 Outbreak: The Impact on Employees' Remote Work Productivity, Engagement, and Stress," *Journal of occupational and environmental medicine* (63:7), p. e426.
Galluch, P. S., Grover, V., and Thatcher, J. B. 2015. "Interrupting the Workplace: Examining Stressors in an Information Technology Context," *Journal of the Association for Information Systems* (16:1), pp. 1-47.
Ganju, K. K., Pavlou, P. A., and Banker, R. D. 2016. "Does Information and Communication Technology Lead to the Well-Being of Nations? A Country-Level Empirical Investigation," *Mis Quarterly* (40:2), pp. 417-+.
Gao, Y., Li, A., Zhu, T., Liu, X., and Liu, X. 2016. "How Smartphone Usage Correlates with Social Anxiety and Loneliness," *PeerJ* (4), p. e2197.
Gerlach, J. P., and Cenfetelli, R. T. 2020. "Constant Checking Is Not Addiction: A Grounded Theory of It-Mediated State-Tracking," *MIS Quarterly* (44:4).
Ghose, A., and Han, S. P. 2011. "An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet," *Management Science* (57:9), pp. 1671-1691.
Griffin, M. A., Neal, A., and Parker, S. K. 2007. "A New Model of Work Role Performance: Positive Behavior in Uncertain and Interdependent Contexts," *Academy of management journal* (50:2), pp. 327-347.
Griffiths, M. 2005. "A 'Components' Model of Addiction within a Biopsychosocial Framework," *Journal of Substance use* (10:4), pp. 191-197.
Haug, S., Castro, R. P., Kwon, M., Filler, A., Kowatsch, T., and Schaub, M. P. 2015. "Smartphone Use and Smartphone Addiction among Young People in Switzerland," *Journal of behavioral addictions* (4:4), pp. 299-307.
Hedges, K. 2014. "Do You Have Fomo: Fear of Missing Out?", from https://www.forbes.com/sites/work-in-progress/2014/03/27/do-you-have-fomo-fear-of-missing-out/

Hobfoll, S. E. 1989. "Conservation of Resources: A New Attempt at Conceptualizing Stress," *American psychologist* (44:3), p. 513.
Horne, J. A., and Ostberg, O. 1976. "A Self-Assessment Questionnaire to Determine Morningness-Eveningness in Human Circadian Rhythms," *International journal of chronobiology* (4:2), pp. 97-110.
Hosio, S., Ferreira, D., Goncalves, J., van Berkel, N., Luo, C., Ahmed, M., Flores, H., and Kostakos, V. 2016. "Monetary Assessment of Battery Life on Smartphones," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1869-1880.
Hülsheger, U. R., Feinholdt, A., and Nübold, A. 2015. "A Low-Dose Mindfulness Intervention and Recovery from Work: Effects on Psychological Detachment, Sleep Quality, and Sleep Duration," *Journal of occupational and organizational psychology* (88:3), pp. 464-489.
Humood, A., Altooq, N., Altamimi, A., Almoosawi, H., Alzafiri, M., Bragazzi, N. L., Husni, M., and Jahrami, H. 2021. "The Prevalence of Nomophobia by Population and by Research Tool: A Systematic Review, Meta-Analysis, and Meta-Regression," *Psych* (3:2), pp. 249-258.
Kang, S., and Jung, J. 2014. "Mobile Communication for Human Needs: A Comparison of Smartphone Use between the Us and Korea," *Computers in Human Behavior* (35), pp. 376-387.
Khaskheli, A., Jun, Y., and Bhuiyan, M. 2017. "M-Commerce and Mobile Apps: Opportunities for Smes in Developing Countries," *Journal of International Business Research and Marketing* (2), pp. 20-23.
Kim, Y., Sohn, D., and Choi, S. M. 2011. "Cultural Difference in Motivations for Using Social Network Sites: A Comparative Study of American and Korean College Students," *Computers in human behavior* (27:1), pp. 365-372.
King, A. L. S., Valença, A. M., and Nardi, A. E. 2010. "Nomophobia: The Mobile Phone in Panic Disorder with Agoraphobia: Reducing Phobias or Worsening of Dependence?," *Cognitive and Behavioral Neurology* (23:1).
King, A. L. S., Valença, A. M., Silva, A. C. O., Baczynski, T., Carvalho, M. R., and Nardi, A. E. 2013. "Nomophobia: Dependency on Virtual Environments or Social Phobia?," *Computers in Human Behavior* (29:1), pp. 140-144.
Király, O., Potenza, M. N., Stein, D. J., King, D. L., Hodgins, D. C., Saunders, J. B., Griffiths, M. D., Gjoneska, B., Billieux, J., and Brand, M. 2020. "Preventing Problematic Internet Use During the Covid-19 Pandemic: Consensus Guidance," *Comprehensive psychiatry* (100), p. 152180.
Konok, V., Gigler, D., Bereczky, B. M., and Miklósi, Á. 2016. "Humans' Attachment to Their Mobile Phones and Its Relationship with Interpersonal Attachment Style," *Computers in Human Behavior* (61), pp. 537-547.
Kuscu, T. D., Gumustas, F., Rodopman Arman, A., and Goksu, M. 2021. "The Relationship between Nomophobia and Psychiatric Symptoms in Adolescents," *International Journal of Psychiatry in Clinical Practice* (25:1), pp. 56-61.
Kuss, D. J., and Griffiths, M. D. 2011. "Online Social Networking and Addiction—a Review of the Psychological Literature," *International journal of environmental research and public health* (8:9), pp. 3528-3552.

Kwon, H. E., So, H., Han, S. P., and Oh, W. 2016. "Excessive Dependence on Mobile Social Apps: A Rational Addiction Perspective," *Information Systems Research* (27:4), pp. 919-939.

Lanaj, K., Johnson, R. E., and Barnes, C. M. 2014. "Beginning the Workday yet Already Depleted? Consequences of Late-Night Smartphone Use and Sleep," *Organizational Behavior and Human Decision Processes* (124:1), pp. 11-23.

Lee, S., Kim, M., Mendoza, J. S., and McDonough, I. M. 2018. "Addicted to Cellphones: Exploring the Psychometric Properties between the Nomophobia Questionnaire and Obsessiveness in College Students," *Heliyon* (4:11), p. e00895.

Lepp, A., Barkley, J. E., and Karpinski, A. C. 2014. "The Relationship between Cell Phone Use, Academic Performance, Anxiety, and Satisfaction with Life in College Students," *Computers in Human Behavior* (31), pp. 343-350.

Ling, R. 2004. *The Mobile Connection: The Cell Phone's Impact on Society | Guide Books | Acm Digital Library*. Morgan Kaufmann Publishers Inc.

Liu, H., Ji, Y., and Dust, S. B. 2021. "“Fully Recharged” Evenings? The Effect of Evening Cyber Leisure on Next-Day Vitality and Performance through Sleep Quantity and Quality, Bedtime Procrastination, and Psychological Detachment, and the Moderating Role of Mindfulness," *Journal of Applied Psychology* (106:7), p. 990.

Magni, M., Ahuja, M. K., and Trombini, C. 2023. "Excessive Mobile Use and Family-Work Conflict: A Resource Drain Theory Approach to Examine Their Effects on Productivity and Well-Being," *Information Systems Research* (34:1), pp. 253-274.

Marengo, D., Sindermann, C., Häckel, D., Settanni, M., Elhai, J. D., and Montag, C. 2020. "The Association between the Big Five Personality Traits and Smartphone Use Disorder: A Meta-Analysis," *Journal of Behavioral Addictions* (9:3), pp. 534-550.

Maslach, C., and Jackson, S. E. 1981. "The Measurement of Experienced Burnout," *Journal of organizational behavior* (2:2), pp. 99-113.

Mattingly, S. M., Gregg, J. M., Audia, P., Bayraktaroglu, A. E., Campbell, A. T., Chawla, N. V., Das Swain, V., De Choudhury, M., D'Mello, S. K., and Dey, A. K. 2019. "The Tesserae Project: Large-Scale, Longitudinal, in Situ, Multimodal Sensing of Information Workers," *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-8.

Mazmanian, M., Orlikowski, W. J., and Yates, J. 2013. "The Autonomy Paradox: The Implications of Mobile Email Devices for Knowledge Professionals," *Organization science* (24:5), pp. 1337-1357.

Melumad, S., and Pham, M. T. 2020. "The Smartphone as a Pacifying Technology," *Journal of Consumer Research* (47:2), pp. 237-255.

Mendoza, J. S., Pody, B. C., Lee, S., Kim, M., and McDonough, I. M. 2018. "The Effect of Cellphones on Attention and Learning: The Influences of Time, Distraction, and Nomophobia," *Computers in Human Behavior* (86), pp. 52-60.

Mi, Z. H., Cao, W. J., Diao, W. J., Wu, M. X., and Fang, X. 2023. "The Relationship between Parental Phubbing and Mobile Phone Addiction in Junior High School Students: A Moderated Mediation Model," *Frontiers in Psychology* (14), p. 10.

Pandey, J., Gupta, M., Behl, A., Pereira, V., Budhwar, P., Varma, A., Hassan, Y., and Kukreja, P. 2021. "Technology-Enabled Knowledge Management for Community Healthcare Workers: The Effects of Knowledge Sharing and Knowledge Hiding," *Journal of Business Research* (135), pp. 787-799.

Park, N., Kim, Y.-C., Shon, H. Y., and Shim, H. 2013. "Factors Influencing Smartphone Use and Dependency in South Korea," *Computers in Human Behavior* (29:4), pp. 1763-1770.
Pelling, E. L., and White, K. M. 2009. "The Theory of Planned Behavior Applied to Young People's Use of Social Networking Web Sites," *Cyberpsychology & behavior* (12:6), pp. 755-759.
Prasad, M., Patthi, B., Singla, A., Gupta, R., Saha, S., Kumar, J. K., Malhi, R., and Pandita, V. 2017. "Nomophobia: A Cross-Sectional Study to Assess Mobile Phone Usage among Dental Students," *J Clin Diagn Res* (11:2), pp. Zc34-zc39.
Ran, W., and Lo, V. H. 2006. "Staying Connected While on the Move: Cell Phone Use and Social Connectedness," *New Media & Society* (8:1), pp. 53-72.
Roodman, D. 2009. "A Note on the Theme of Too Many Instruments*," *Oxford Bulletin of Economics and Statistics* (71:1), pp. 135-158.
Rubino, C., Perry, S. J., Milam, A. C., Spitzmueller, C., and Zapf, D. 2012. "Demand-Control-Person: Integrating the Demand-Control and Conservation of Resources Models to Test an Expanded Stressor-Strain Model," *J Occup Health Psychol* (17:4), pp. 456-472.
Rutkowski, A.-F., and Saunders, C. 2018. *Emotional and Cognitive Overload: The Dark Side of Information Technology*. Routledge.
Samaha, M., and Hawi, N. S. 2016. "Relationships among Smartphone Addiction, Stress, Academic Performance, and Satisfaction with Life," *Computers in Human Behavior* (57), pp. 321-325.
Santoro, G., Vrontis, D., Thrassou, A., and Dezi, L. 2018. "The Internet of Things: Building a Knowledge Management System for Open Innovation and Knowledge Management Capacity," *Technological forecasting and social change* (136), pp. 347-354.
Sarker, S., Campbell, D. E., Ondrus, J., and Valacich, J. S. 2010. "Mapping the Need for Mobile Collaboration Technologies: A Fit Perspective," *International Journal of e-Collaboration (IJeC)* (6:4), pp. 32-53.
Sarker, S., Sarker, S., Xiao, X., and Ahuja, M. 2012. "Managing Employees' Use of Mobile Technologies to Minimize Work-Life Balance Impacts,").
SecurEnvoy. 2012. "66% of the Population Suffer from Nomophobia the Fear of Being without Their Phone." from https://securenvoy.com/blog/66-population-suffer-nomophobia-fear-being-without-their-phone-2/
Serenko, A., Bontis, N., and Hull, E. 2016. "An Application of the Knowledge Management Maturity Model: The Case of Credit Unions," *Knowledge Management Research & Practice* (14:3), pp. 338-352.
Sharma, M., Amandeep, Mathur, D. M., and Jeenger, J. 2019. "Nomophobia and Its Relationship with Depression, Anxiety, and Quality of Life in Adolescents," *Ind Psychiatry J* (28:2), pp. 231-236.
Shipley, W., Gruber, C., Martin, T., and Klein, A. 2009. "Manual Shipley-2," *Los Angeles: Western psychological services*).
Soror, A. A., Hammer, B. I., Steelman, Z. R., Davis, F. D., and Limayem, M. M. 2015. "Good Habits Gone Bad: Explaining Negative Consequences Associated with the Use of Mobile Phones from a Dual-Systems Perspective," *Information Systems Journal* (25:4), pp. 403-427.
Soto, C. J., and John, O. P. 2017. "The Next Big Five Inventory (Bfi-2): Developing and Assessing a Hierarchical Model with 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power," *Journal of personality and social psychology* (113:1), p. 117.

Spielberger, C. D. 1983. "State-Trait Anxiety Inventory for Adults,").
Srinivasan, K., Currim, F., and Ram, S. 2023. "A Human-in-the-Loop Segmented Mixed-Effects Modeling Method for Analyzing Wearables Data," *ACM Trans. Manage. Inf. Syst.* (14:2), p. Article 18.
Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., and Bühner, M. 2020. "Predicting Personality from Patterns of Behavior Collected with Smartphones," *Proceedings of the National Academy of Sciences* (117:30), pp. 17680-17687.
Tafti, A., and Shmueli, G. 2020. "Beyond Overall Treatment Effects: Leveraging Covariates in Randomized Experiments Guided by Causal Structure," *Information Systems Research* (31:4), pp. 1183-1199.
Takao, M., Takahashi, S., and Kitamura, M. 2009. "Addictive Personality and Problematic Mobile Phone Use," *CyberPsychology & Behavior* (12:5), pp. 501-507.
Tams, S., Legoux, R., and Léger, P.-M. 2018. "Smartphone Withdrawal Creates Stress: A Moderated Mediation Model of Nomophobia, Social Threat, and Phone Withdrawal Context," *Computers in Human Behavior* (81), pp. 1-9.
Thakur, A., Gormish, M., and Erol, B. 2011. "Mobile Phones and Information Capture in the Workplace," *CHI '11 Extended Abstracts on Human Factors in Computing Systems*: ACM.
Turel, O., and Serenko, A. 2010. "Is Mobile Email Addiction Overlooked?," *Communications of the ACM* (53:5), pp. 41-43.
van Deursen, A. J. A. M., Bolle, C. L., Hegner, S. M., and Kommers, P. A. M. 2015. "Modeling Habitual and Addictive Smartphone Behavior: The Role of Smartphone Usage Types, Emotional Intelligence, Social Stress, Self-Regulation, Age, and Gender," *Computers in Human Behavior* (45), pp. 411-420.
Wang, G., Suh, A., and Acm. 2018. "Disorder or Driver?: The Effects of Nomophobia on Work-Related Outcomes in Organizations," *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems (Chi 2018)*), p. 12.
Watson, D., and Clark, L. A. 1994. "The Panas-X: Manual for the Positive and Negative Affect Schedule-Expanded Form,").
Williams, L. J., and Anderson, S. E. 1991. "Job Satisfaction and Organizational Commitment as Predictors of Organizational Citizenship and in-Role Behaviors," *Journal of management* (17:3), pp. 601-617.
Yildirim, C., and Correia, A. P. 2015. "Exploring the Dimensions of Nomophobia: Development and Validation of a Self-Reported Questionnaire," *Computers in Human Behavior* (49), pp. 130-137.
Yu, L., Cao, X., Liu, Z., and Wang, J. 2018. "Excessive Social Media Use at Work: Exploring the Effects of Social Media Overload on Job Performance," *Information technology & people* (31:6), pp. 1091-1112.
Zhai, X., Wang, M., Chen, N.-S., Ghani, U., and Cacciolatti, L. 2023. "The Secret Thoughts of Social Network Sites Users: A Scale for the Measurement of Online Knowledge-Hiding in a Knowledge Exchange (Ke) Context," *Interactive Learning Environments* (31:5), pp. 2899-2913.
Zhang, Z., and Ji, X. 2022. "A Virtual Net Locks Me In: How and When Information and Communication Technology Use Intensity Leads to Knowledge Hiding," *Journal of Business Ethics*), pp. 1-16.

Conclusions

This dissertation examines the complex interplay between data, algorithms, and human behavior. We explore the ethical and behavioral complications that arise from the prevalence of online data collection in light of privacy norms, the replacement of human experts with algorithms, and the proliferation of mobile phone use in the workplace. While seemingly distinct ideas, the underlying phenomena to each of these ideas is shared: that individuals learn and adapt to interact with advancing technologies. As decision makers, individuals navigate the technological landscape every day and are constantly being asked to update their knowledge and make wise decisions. This work highlights the complexities that arise from that ask.

Chapter 1 starts by considering data privacy and online decision making. With a novel experimental design, we ask participants to make a privacy decision. We vary the presentation of this decision in line with the requirements of the General Data Protection Regulation. We then manipulate the inherent riskiness of the privacy choice and examine if consumers are able to adequately adjust their risk profile in light of this change. We find that the enhanced privacy controls set forth by GDPR are more effective at changing rates of consent in low-risk, rather than high-risk, settings. This is a concerning result that highlights the need for higher precision in the targeting of privacy regulation to ensure a balance between the flourishing of the data economy and the privacy rights of individuals.

Chapter 2 examines the role of betrayal aversion in the adherence to expert advice. Betrayal aversion, defined as the strong dislike for the violation of trust norms, has been well-studied in economics but had not yet been introduced to the IS literature. We develop an intricate simulated financial market and give participants the opportunity to invest real money in the

market. We give them access to either a human or algorithmic expert to aid them in their decision making. The risk of betrayal is then manipulated and we examine how it subsequently impacts the participants' willingness to utilize the expert advice. We find that betrayal aversion decreases utilization by 16% in the human advisor condition. However, that effect is entirely attenuated when the human is replaced with an algorithm. This work has important implications for how humans apply social rules and norms to algorithms.

Finally, Chapter 3 explores the role of nomophobia, or the fear of being without one's mobile phone, in the workplace. We use a unique longitudinal data set collected from the Tesserae Project. This was a year long study of 757 information workers across the US who wore health wearables, placed location beacons in their home and at their work, download a mobile phone app that could track usage, and respond to daily surveys. Utilizing that data, we identify the causal effect of decreasing phone-battery level (our operationalization of nomophobia) on employee stress. We find that decreasing battery does in fact increase stress and this indirectly decreases productivity measures as well. However, in follow-up analysis, we show the moderating effect of remote work. We find that working from home is able to attenuate the productivity declines brought about by nomophobia.