Fall 12-2020

The Effects of Security Framing, Time Pressure, and Brand Familiarity on Risky Mobile Application Downloads

Cody Parker

Old Dominion University, codyparker.ux@gmail.com

THE EFFECTS OF SECURITY FRAMING, TIME PRESSURE, AND BRAND FAMILIARITY ON RISKY MOBILE APPLICATION DOWNLOADS

by

Cody Parker
B.A. May 2016, University of Central Oklahoma

A Thesis Proposal Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December 2020

Approved by:

Jing Chen (Director)

Jeremiah Still (Member)

Abby Braitman (Member)

ABSTRACT

THE EFFECTS OF SECURITY FRAMING, TIME PRESSURE, AND BRAND FAMILIARITY ON RISKY MOBILE APPLICATION DOWNLOADS

Cody Parker
Old Dominion University, 2020
Director: Dr. Jing Chen

The current study examined the effects of security system framing, time pressure, and brand familiarity on mobile application download behaviors, with an emphasis on risk taking. According to the Prospect Theory, people tend to engage in irrational decision making, and make qualitatively different decisions when information is framed in terms of gains and losses (i.e., the framing effect). Past research has used this framing effect to guide the design of a risk display for mobile applications (apps), with the purpose of communicating the potential risks and minimizing insecure app selections. Time pressure has been shown to influence the framing effect in both hypothetical choices in lab settings as well as with consumer purchases, and brand familiarity has been shown to affect consumers' purchase behaviors. Neither factor has been studied in the context of risk communication for mobile app. The current study addressed this gap in the literature and examined the effects of time pressure and brand familiarity on the effectiveness of risk displays (framed as safety or risks) for mobile apps. Specifically, users' choices were recorded as a measure of effective risk displays. The findings from this study indicated that users rely heavily on brand familiarity when downloading apps. We also showed that security scores, especially when framed as safety, were effective at guiding choice, though this advantage of safety framing was not present when users made decisions under time pressure. The implications from the study indicate that people implicitly trust brands they recognize, safety framed security can be helpful, and decision-making processes change under time pressure.

This thesis is dedicated to my best friend and wife, Alyssa.

ACKNOWLEDGMENTS

I would like to thank everyone whose hands have helped shaped this thesis. I would like to specifically thank my advisor and committee chair, Dr. Jing Chen, for the hours and effort she has dedicated guiding me through this project. I would also like to thank my committee members, Dr. Jeremiah Still and Dr. Abby Braitman, for their commitment to this thesis.

NOMENCLANTURE

Apps: Applications

GLMER: General Linear Mixed-Effects Regression

LRT: Likelihood Ratio Test

HLM: Hierarchical Linear Modeling

ICC: Intra-Class Correlation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

While smartphones have become commonplace and are treated differently than desktop computers (McGill & Thompson, 2017), this does not make them inherently free from cybersecurity threats. The ubiquity of smartphones and their contained applications (apps) cannot be disputed, with an estimated 3.3 billion smartphone users as of 2019 (Holst, 2019) and an estimated 2.7 million android apps available as of June 2019 (Clement, 2019). Unfortunately, not all apps are benign; there are those that seek to gather user data through illegal or unethical practices. For example, some apps may access bank passwords by “sniffing” the magnetometer and accelerometer (Ning et al., 2018), and other apps, like Facebook, use complex legalese and numerous permissions to gather and sell contact lists, location, browsing history, and other data from users (Jaeger, 2014). Both types of intrusions embody predatory practices against users, many of whom unknowingly and unnecessarily subject themselves to cybersecurity threats (Price, 2018). For example, in October 2019, 42 Android apps were found to contain malware, having been downloaded over 8 million times by users (Stefanko, 2019). Because of imperfect screening of apps and a lack of consumer scrutiny, malicious developers are capable of victimizing users (Price, 2018). It is therefore of great importance that users make safe, informed decisions regarding the apps they choose to download onto their devices. These secure, informed decisions rely on effective communication of the potential risks associated with mobile apps to the users (Chen, Gates, Li, & Proctor, 2015).

The process of consumer decision making in general can be influenced by external factors, some of which haven shown to result in irrational or unsafe practices. These external factors include how the information is framed (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981), time pressure (Young, Goodie, Hall, & Wu, 2012; Saqib & Chan, 2015), misplaced trust in brand familiarity (Baker, Hutchinson, Moore, & Nedungadi, 1986), and so on. However, these factors have not been investigated extensively for mobile app downloads behaviors specifically. The current study aims to examine these factors for mobile app selections to inform design tools that can be implemented to combat such factors. The findings from the study are expected to guide better app risk display design that puts the privacy and well-being of users first.

1.1 Importance of Risk Communication for Mobile Applications

In order to function properly, apps may need access to various sensors and folders of information within a device, such as location, contacts, and camera (Felt et al., 2012). However, before digging through a user's phone, apps must first ask the user's permission for the different sources of information. While certain permissions, such as location for GPS navigation apps, are legitimate when required for the app to function, apps may not always request access to information directly related to its function. Instead, an app may gather completely unrelated information and sell or use the data for other motives (Vidas, Christin, & Cranor, 2011). Moreover, an app may gather such information when the app is not actively in use (Nakashima, 2018).

Given the gravity of such data abuse, it is concerning that smartphone users tend not to investigate or fully understand permissions requested by the apps when downloading new apps (Chin, Felt, Sekar, & Wagner, 2012; Felt et al., 2012; Kelley et al., 2012; Benton, Camp, & Garg, 2013). This lack of permissions understanding makes it difficult for the average user to fully grasp the potential threats associated with downloading an app. After all, apps are currently framed in the best lighting, clearly showing the benefits associated with the app, whether that is the entertainment of a game, the functionality of a utility app, or the ability to connect with others through social media. The potential losses of privacy or sensitive data, like passwords or banking information, stays relatively latent unless actively sought out. If users wish to prioritize safe practices, the burden is put on them to understand permissions and closely read terms of agreement in order to make informed choices, but this is rarely the approach people take (Felt et al., 2012). Not surprisingly, when the act of investigating potential losses requires great effort, people tend to be more *risk taking*, choosing not to engage in extra work (Maule, Hockey, & Bdzola, 2000). Indeed, by designing app stores to highlight benefits and suppress losses, users are not expected to actively search for potential threats. Fortunately, the Android system changed permission requests when it updated from version five (“lollipop”) to version six (“marshmallow”) in both presentation and allowance (Moore, Ge, Li, & Proctor, 2019). Whereas the fifth version required users to allow apps to access all permissions, version six allowed users to choose which permissions were accessed. Moore et al. showed that users both preferred this format and better understood the functions of the permission requests. However, version six was not as effective at conveying all of the permission requests to users and was not found to be more effective at reducing risky engagement.

Another distressing trend is for users to approach app purchases with a trial basis mentality, deciding to keep or delete the app only after using it (Kim, Kankanhalli, & Lee, 2016). Tragically, once the app is downloaded, the developers have immediate access to personal information. By the time a user may decide to delete the app, the developer may have already mined their data. Therefore, interventions should strive to inform users of potential risks before downloading the app in question. One way to educate and empower users is to provide them with simple, explicit displays of risks for apps that appear to be malicious (Choe, Jung, Lee, & Fisher, 2013; Gates, Chen, Li, & Proctor, 2014; Chen et al., 2015). These studies have proposed and refined the design of security scores that summarizes the safety or risk associated with an app, based its permission requests. This approach is expected to reduce risky, uneducated downloads and increase safer choices. The proposed security system, as briefly mentioned, would analyze the permission requests of apps in order to display security ratings. Research by Peng et al. (2012) indicated that malicious apps contain more permission requests than their counterparts; based on this, they used machine learning and Naïve Bayes to predict malicious apps in the Google Play Store. Though there is currently no system like this in either the Google Play Store or the App Store, there is a security suite by Appthority that examines application safety, similar to the security system proposed. Before being acquired by Symantec (Sawers, 2018), Appthority was a company that analyzed developers' data mining practices, app permissions, and mobile traffic patterns to dig into what an app did with a user's data; Symantec now provides this service for companies and their employees. For real-world implementation, the risk displays for mobile apps would similarly analyze the permissions of an app, its functionality, behavior, and data collection to provide a security score within the mobile app stores (e.g., Google Play Store and App Store) for all users, because all mobile users should be provided the opportunity to safely use their devices. However, this begs the question of how to effectively display the calculated risk score to promote secure app-selection decisions of common users. In other words, the framing of such information is of vital importance. For example, whether the system should utilize a safety score (the greater the score, the safer the app) or a risk score (the greater the score, the more dangerous the app; Chen et al., 2015; Chong, Ge, Li, & Proctor, 2018). These studies have proposed and refined the design of security scores that summarizes the safety or risk associated with an app, based its permission requests. This approach is expected to reduce risky, uneducated downloads and increase safer choices. The proposed security system, as briefly mentioned, would analyze the permission requests of apps in order to display security ratings.

1.2 Prior Studies on Risk Displays for Mobile Applications

Past research on the risk displays for mobile apps has suggested that the framing of the risk information can influence users' app-selection behaviors (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018). For example, Chen et al. (2015) included a summary risk/safety rating for each app and showed a positive effect of the risk/safety rating on app selection, although there are other factors that have been shown to influence app selection as well. Indeed, Rajivan and Camp (2016) and Chong et al. (2018) found that, beyond positively framing a security system with safety scores, users can be influenced to make safer decisions when primed. Furthermore, Rajivan and Camp tested the effectiveness of different iconography on promoting safe decision making and found that locks resulted in the safest choices, likely due to the familiarity of the icon and associated mental models with web browsers. Alternatively, Chen et al. (2018) investigated intermediate-level risk displays containing three major risk categories as opposed to the summary risk/safety ratings in the research by Chen et al. (2015), Rajivan and Camp (2016), and Chong et al. (2018), and found that displaying these risk categories promoted secure app selections and allowed for potential personalization of risk displays based on the user's risk concerns. Finally, a study by Shuster et al. (2015) tested users' preference for risk presentation between automated reports and human reviews. The authors displayed apps in a semi-realistic application store environment and presented security categorically along a scale of potential danger ("safety", "caution", and "risky") instead of using framed presentations. That is, participants were not presented with either "safety" or "risky" but were shown all three categories to emulate a range of categories instead of shown security along a framed scale of safety or riskiness, respectively. They found that participants were more likely to choose safer apps than risky apps and were more trusting of automated reports than of human reviews, with males in particular being more trusting of automation than females.

Chen et al. (2015) conducted three experiments to gauge the effectiveness of a summary security score and how the framing of such a score influenced app choice. In all three of Chen et al.' experiments, the independent variables were user ratings (out of five stars), security score (out of five circles), and security type (safety, risk). The dependent variables were whether or not an app was chosen out of six apps in a given trial, and the subjective rationale for the apps chosen. Participants were shown a list of six apps per trial, each with a security score (framed as either risk or safety), user ratings (scores and count of users), icons, and brief descriptions provided by the developers. Upon choosing two apps per trial, dropdown menus below the chosen apps listed rationale choices from which participants could choose as many as were applicable. The rationale choices were as follows: "User Rating Score, User Rating Count, Permission Safety (Or Risk), Icon Look and Feel, Description, Familiarity with app or developer, Other" (Chen et al., 2015; p. 152).

The choice data in each of Chen et al.'s (2015) experiments were analyzed using a logistic regression and a repeated-measures ANOVA while the rationale data were analyzed using Chi square analyses. When the system was framed as app safety as opposed to risk, risk taking behaviors were significantly reduced for all three experiments, though both framed systems decreased risk taking compared to no security system (Chen et al., 2015). In other words, participants avoided the apps with low safety ratings, choosing alternative apps instead.

Research by Rajivan and Camp (2016) further examined the effect of security priming on app selection. The experiment was designed to simulate an android phone and participants were prompted to choose four of the eight presented apps. The procedure had participants go through multiple steps, comparing the apps against each other, examining app information, and “downloading” apps and seeing the permission requests. Additionally, they introduced framed icons (frowning emoji, a red eye, or a lock) instead of the color-coded scale used by Chen et al. (2015) to indicate the riskiness of apps. The authors found that the positively framed lock was more effective than the negatively framed frowning emoji and the red eye. They also showed that priming was only effective at influencing the selection of the first app and significantly increased decision times.

Chong et al. (2018) provided further validation for the findings from Chen et al.’s (2015) and Rajivan and Camp’s (2016) studies. Chong et al. showed that priming can be effective, whether self-relevant or factual. Chong et al.’s experimental design followed that of Chen et al.’s study and expanded on it by investigating the effect of security priming. Furthermore, Chong et al. showed that the positively framed safety scores continued to be more effective than risk scores at guiding users to make safer choices. Importantly, Chong et al. found that priming can increase safe behavior compared to no priming. The authors examined the length of the priming process (priming with eight example apps compared to priming with one app) and found that priming with just one app can increase safe behaviors.

Ideally, the proposed risk displays would make it easier for users to understand whether an app was relatively safe without having to dig into specific details or reasons. Therefore, the inclusion of such a tool is anticipated to reduce risk-taking behaviors by bringing potential threat information to the users’ attention. The past research (Gates et al., 2014; Chen et al., 2015;

Rajivan & Camp, 2016; Chong et al., 2018) have advanced the understanding of security framing. The current study serves to further test the effectiveness of security score framing in the context of external factors that could influence this framing effect but have not been considered by the previous experiments. This study is the first to consider time pressure and brand familiarity. Indeed, all of the previously listed literature controlled for brand familiarity by removing the top search results from inclusion in the experimental stimuli, and had participants perform the task using as much time as they needed (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018). However, it is realistic that people will install apps under time pressure and that they may be influenced by brand familiarity. For example, one type of time pressure that mobile users may encounter is when situations suddenly require a specific app, such as when a user may need to download a money-sharing app on the spot in order to pay a debt, or when someone wants to rent a shared electric scooter for faster travel. Additionally, users may be more likely to choose an app they recognize under time pressure; for example, if someone needed to download a PDF reader app, he or she may choose one by Adobe due to its familiarity. If time pressure impacts the decision-making strategies of mobile users when choosing apps, the proposed security scores may not be universally effective under the safety (gain) frame, thus requiring further examination.

1.3 The Framing Effect

Decision making research has repeatedly shown that, as irrational beings, humans are inherently prone to error (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981). According to Kahneman and Tversky's Prospect Theory (1979), decision makers do not treat gains and losses equally. Figure 1 shows an S-curve, highlighting the differences between subjective and objective values, as predicted by Prospect Theory.

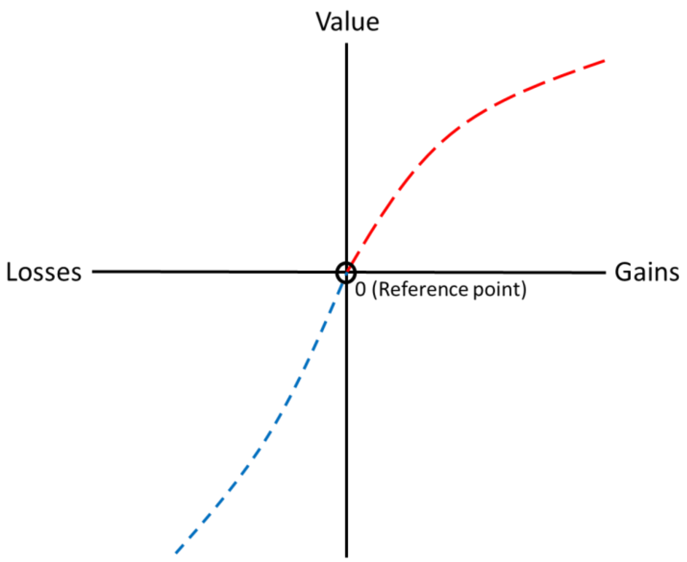


Fig. 1. The S-curve, as predicted by Prospect Theory (Tversky & Kahneman, 1981). On the X-axis are objective gains and losses, with greater distance from the status quo (reference point) indicating greater gains and losses, respectively. On the Y-axis are the subjective values assigned to the gains and losses. The red dashed line represents risk aversion while the blue dotted line represents risk seeking.

As is clearly seen with the curvilinear lines, gains and losses are not assigned rational values by people. Where a straight line would indicate subjective values that are proportional to objective values, the curvilinear lines suggest that subjective value differences are much greater when closer to the reference point. In other words, an objective value, such as five dollars, can be perceived very differently depending on the context of the question. For example, the subjective difference between $5 and $10 is the much greater than the difference between $95 and $100. The S-curve reflects higher subjective values with steeper lines near the reference point (the origin). Furthermore, the curve for losses is steeper, indicating greater subjective value for losses than for objectively equal gains; a loss of $5 would have a stronger subjective value than a gain of $5. The Prospect Theory thus shows that people approach equitable choices differently based on the way in which they are framed, whether it leads to gains or losses. This can be seen most clearly in the experimental questions created for the framing effect, wherein participants are given the option to choose a sure option or a risky option that contained a gamble (see Table 1).

Table 1

An example question from Fagley and Kruger's (1986) on the framing effect.

Example Question from Fagley and Kruger (1986)			
Imagine that in one particular state it is projected that 1,000 students will drop out of school during the next year. Two programs have been proposed to address this problem, but only one can be implemented. Based on other states' experiences with the programs, estimates of the outcomes that can be expected from each program can be made. Assume for purposes of this decision that these estimates of the outcomes are accurate and are as follows:			
Positive/Gain		Negative/Loss	
	If program 1 is adopted, 400 of the 1,000 students will stay in school.		If program 1 is adopted, 600 of the 1,000 students will drop out of school.
	If program 2 is adopted, there is 2/5 chance that all 1,000 students will stay in school and 3/5 chance that none of the 1,000 will stay in school.		If program 2 is adopted, there is 2/5 chance that none of the 1,000 will drop out of school and 3/5 chance that all 1,000 students will drop out of school.

In the example question, both frames are shown, with the gain frame emphasizing student retention and the loss frame emphasizing dropout rates. The retention and dropout rates are equal for all programs, with one option per frame being certain, or "sure," and the other per frame being a gamble, or "risky." If people were rational, there should not be a difference in preference for either the risky or certain option. However, due to the subjective values assigned according to the Prospect Theory (Tversky & Kahneman, 1981), Fagley and Kruger (1986) showed that, when choices are framed as gains (retaining students), people are typically risk averse (choosing to keep a set amount of students), whereas choices that are framed as potential losses (student dropout) result in risk taking behaviors (gambling in order to eliminate potential dropout). In other words, people tend to choose certain gains and are willing to "risk it all" for a chance to mitigate whatever losses they can. Therefore, risk aversion is the desire to make a safe, sure choice, while risk-seeking is less concerned with safety, but are willing to gamble with the possibility of either gaining nothing (for gain frames) or losing everything (for loss frames). Unfortunately, this framing effect is not limited just to novice participants in laboratory settings.

Even experts in their respective fields are prone to being influenced into predictable patterns of choice. For example, medical professionals have been shown to make more conservative medical decisions under gain frames and choose more aggressive procedures under loss frames (Mazur & Hickam, 1990). Likewise, national intelligence analysts are susceptible to gambling hypothetical human lives under loss frames compared to gain frames and expressed greater confidence in their decisions compared to college students (Reyna, Chick, Corbin, & Hsia, 2014). Therefore, depending on the way in which their respective decisions are framed, these professionals may take unnecessary risks. Indeed, Reyna et al. (2014) showed that professionals can be more likely to make erroneous decisions with greater commitment than the general population.

The framing effect is thus pervasive and can be employed to influence decision making. Therefore, designers can make use of this principle to increase or decrease user risk taking, according to the designer's goals. In the domain of mobile cybersecurity, the framing effect can be implemented in security scores for mobile app safety or risk. The composite security score first proposed by Gates et al. (2014) and further expanded upon by Chen et al. (2015) used the framing effect in their design of such a system.

1.4 The Effects of Time Pressure on Decision Making

Time pressure, whether explicit or perceived, induces a sense of urgency when attempting a task (Klapproth, 2008) and can affect the way people make decisions (Young et al., 2012). In consumer decision-making literature, there are a number of findings that indicate the substantial influence that time pressure, whether real or contrived, can have on shopping behaviors and perceptions. Wright (1974) is often cited for his foundational research on consumer behavior under time pressure. His study shows that, under time pressure, consumers tend to emphasize negative traits of a product, indicating a change in decision making processes. Since then, a plethora of research has been conducted on decision making and risk taking in various shopping environments, suggesting that time pressure, via scarcity of products (Devlin, Ennew, McKechnie, & Smith, 2007; Soliman, 2017) or length of sale (Aggarwal & Vaidyanathan, 2003), can dictate the strategy with which consumers approach purchases (Chowdhury, Ratneshwar, & Mohanty, 2009; Vlašić, Janković, & Kramo-Čaluk, 2011) and their acceptance of risk (Shehryar, 2008).

Outside the niche domain of consumer choices under time pressure, the more general decision-making literature has examined the effect of time pressure on risk taking for decades, with a plethora of competing findings. For example, Ben Zur and Breznitz (1981) showed that, under time pressure, participants were less likely to take risks with hypothetical gambles that contained the possibility of both gain and loss and were more likely to focus on the negative aspects of each gamble. By including both gains and losses in each trial, the study by Ben Zur and Breznitz did not follow the standard structure to examine the framing effect; therefore, the global reduction of risk taking cannot be directly applied to such research. Furthermore, El Haji, Krawczyk, Sylwestrzak, and Zawojska (2016) showed that, under time pressure (25 seconds versus six minutes), people are less likely to bid on a lottery, suggesting that people are more risk averse than without time pressure. This global reduction in risk taking under time pressure may be attributed to the fact that participants had to bid on the lottery, thus involving a potential loss rather than a net zero outcome; in this way, both the potential to gain and lose money makes this study difficult to directly compare to those that invoke the framing effect.

Contrary to the literature that suggests an overall reduction in risk taking under time pressure, Chandler and Pronin (2012) found that, after being prompted to read sentences at a fast pace, participants were more willing to take risks when completing the Balloon Analogue Risk Task (BART) than their slow-paced counterparts. In this task, participants would inflate digital balloons for a monetary reward; however, if they popped the balloons, they forfeited the reward. Likewise, Madan, Spetch, and Ludvig (2015) showed a greater risk acceptance for participants under time pressure with a gambling task. The task was for participants to choose between two color-coded doors for any given trial, with one door providing a moderate, unchanging reward and the other providing either a great reward or nothing, much like the positively framed questions created by Tversky and Kahneman (1981).

These competing results seem to be due to the different frames used in the studies. The research by Ben Zur and Breznitz (1981) and El Haji et al. (2016) suggest that people become risk averse under time pressure when faced with both potential gains and losses within a given decision. Alternatively, research that utilized potential gains, as in the studies by Chandler and Pronin (2012) and Madan et al. (2015), suggest that people are more risk seeking under time pressure. According to the framing effect, it is unlikely that people approach both gains and losses with the same strategies. Fortunately, there is literature that takes the framing effect into consideration when studying the effects of time pressure on decision making.

Instead of finding overall risk aversion or seeking behaviors without considering the framing of the question, Young et al. (2012) found that gain frames lead to riskier gambles under time pressure, a finding consistent with Chandler and Pronin’s (2012) and Madan et al.’s (2015) results. Additionally, Young et al. (2012) showed extreme risk taking for loss frames under time pressure and that participants became worse at approximating the true values of probabilities. The authors attributed this extreme risk taking for loss frames to the inaccurate estimations; as such, the findings required further investigation. Based partly on the findings from Young et al. (2012), Saqib and Chan (2015) proposed that people tend to focus on the maximal possible values under time pressure instead of the status quo. However, contrary to Young et al.'s (2012) findings, Saqib and Chan (2015) suggested an inversion of risk preferences for gains and losses under time pressure.

For decisions without time pressure, the framing effect suggests that people use the status quo, not present losses or gains, as a reference point. However, under time pressure, Saqib and Chan posit that decision makers fixate on the maximal values (gains or losses, depending on the frame) and use these values as the reference point instead of the status quo. Therefore, any gains or losses that are less than these maximal values are now perceived as being losses and gains, respectively. As a result, people are more risk taking with gain frames and more risk averse with loss frames under time pressure than without time pressure. Using the example question from Fagley and Kruger's (1986) study, people under time pressure are more likely to focus on the maximal value of 1,000 students instead of 400 dropouts or 600 retained. Because of the fixation on maximal values under time pressure, the comparative difference between the smaller "sure" options and the larger "risky" option are perceived as losses in gain frames and gains in loss frames. Figure 2 shows this inversion of the Prospect Theory's S-curve.

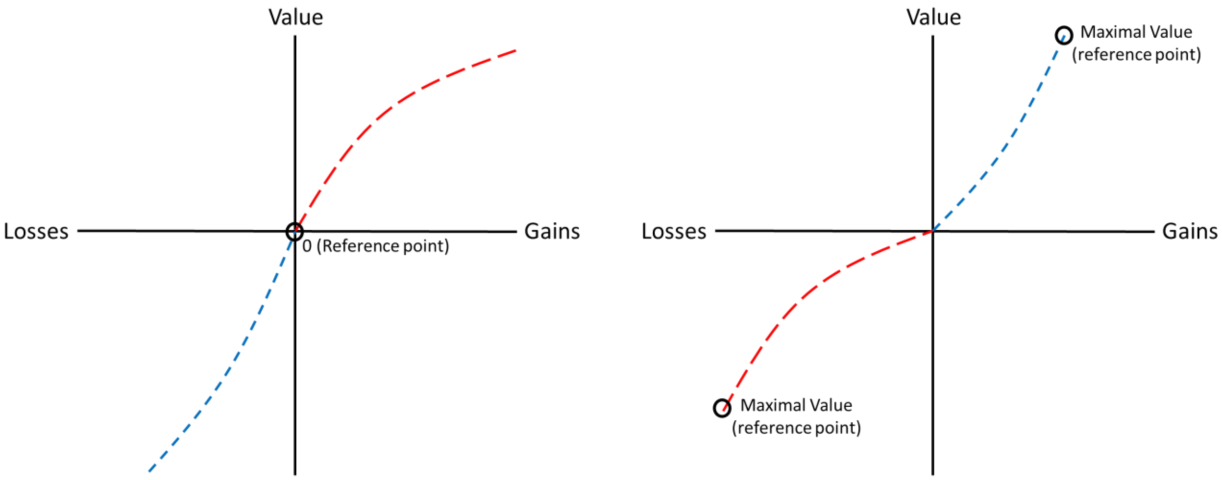


Fig. 2. Hypothetical inversion of Prospect Theory's S-curve, as proposed by Saqib and Chan (2015). On the left is the normal S-curve, while the S-curve on the right is under time pressure. The inversion is based on the idea that, with maximal values acting as the new reference point, the curves that were previously associated with gain and loss frames flip. The maximal value reference points essentially lead to people treating sure gains (losses) as losses (gains).

It should be noted that this finding by Saqib and Chan (2015) is relatively recent and requires further validation; a study by Wegier and Spanniol (2015) found a similar inversion of risk preferences under time pressure, but their results showed a significant decrease in risk seeking with the loss frame and a non-significant trend of increase in risk seeking with the grain frame. Note that this finding does not reflect that by Saqib and Chan (2015) or that by Young et al. (2012). It is therefore clear that the effects of time pressure on decision making are not fully understood, given the vastly different findings by various researchers. As a result, further research is needed to understand the influence of time pressure on risky decisions for decisions made under gain and loss frames.

1.5 Effect of Brand Familiarity on Consumer Behavior

As previously mentioned, threats to mobile users can come in a variety of forms, from a variety of sources. While some lesser known developers have been caught writing malware into their apps (Price, 2018), more familiar developers may choose to collect and sell user data to third parties (Wong, 2019). Both forms of exploitation are cause for concern but may be represented differently in the minds of users, with more familiar apps seeming more innocuous (Harris, Brookshire, & Chin, 2016). Indeed, according to Baker et al. (1986), brand familiarity both increases positive affect and drives purchase behaviors. For the study at hand, we define brand familiarity as recognition of a company and/or its product; familiarity is not the same as favorability. In this study, people can be familiar with a company or product but dislike it.

There has been little research that examines the impact of brand familiarity on purchase intentions for mobile apps. The effect of brand familiarity is especially important in light of the allegations against companies like Facebook, which was recently fined five billion dollars for its recklessness with user data (Wong, 2019). Despite the general population's outcries against the mishandling of data, there has been a linear increase in Facebook users since 2008 (Clement, 2020). Therefore, although brand familiarity is likely to positively influence product purchase behaviors in more general domains, it is important to the understanding of mobile security that the effect of brand familiarity be studied. If the effect of brand familiarity holds as strongly for mobile apps as with more general products, both researchers and security score designers should be aware of its impact and how they can overcome it.

To date, though, no studies have examined the effect of brand familiarity on the proposed security system. Indeed, Chen et al. (2015) and Chong et al. (2018) controlled for brand familiarity in their study by skipping the top ten apps in the search results of a given app functionality, while Rajivan and Camp (2016) skipped the top 75 apps. To date, there is no research that examines the interactive effect of brand familiarity and security scores on app selection. However, previous literature shows a strong effect of brand familiarity on purchase intentions in more general consumer domains, such as apparel (Park & Stoel, 2005) and cold medicines (Laroche, Kim, & Zhou, 1996). Therefore, because of the effect brand familiarity has on purchase intentions, it is expected that brand familiarity may be more influential than security scores in guiding download behaviors. Indeed, this effect is expected to be particularly strong for apps that lack true alternative options, such as social media platforms.

1.6 Current Study

As discussed above, there are several factors that play into a consumer's decision to engage with a brand, let alone download an app onto their personal device. In the real world, factors within the app store like user ratings (Zhu & Zhang, 2010) and brand familiarity (Harris et al., 2016) are expected to play a role in app selection. Likewise, external factors such as personality (Xu, Frey, Vuckovac, & Ilic, 2015), social influence (Zhu & Zhang, 2010), and time pressure (Young et al., 2012; Saqib & Chan, 2015) are expected to contribute to the decision-making process. However, while these (and other) factors are likely to guide consumer behaviors, the current study focused on time pressure and brand familiarity.

Given the past research on decision making and risk taking (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981; Young et al., 2012; Saqib & Chan, 2015), particularly under time pressure (Young et al., 2012; Saqib & Chan, 2015), it stands to reason that mobile users would engage in even riskier download behaviors under situational time constraints. Because of this preference reversal, there is reason to believe that the support for security scores by Chen et al. (2015) may differ under varying conditions, as with perceived time pressure. Indeed, time pressure has been shown to impact the ability of consumers to investigate product information, making it difficult for consumers to notice when brands attempt to omit negative product information (Kardes et al., 2006; do Prado & Lopes, 2016). Furthermore, research by Liu, Hsieh, Lo, and Hwang (2017) showed that the effect of brand familiarity on browsing behaviors can be influenced by time pressure, with consumers spending more time and looking more often at recognizable brands than less familiar alternatives under time pressure than without time pressure. Without time pressure, participants viewed both familiar and unfamiliar brands equally, though the authors did not report any findings on product choice, instead relying solely on eye-tracking data. This proxy for choice is acknowledged by the current study and is used to guide app choice hypotheses for the main experiment.

The goals of the current research were to study the effects of brand familiarity, security scores, and perceived time pressure on mobile app download behaviors in order to better understand the factors that increase risky downloads as well as the best way to reduce these effects. We conducted a pilot study to measure brand familiarity and trustworthiness for apps and to examine the correlation between these constructs. Afterwards, an experiment examined the effects of security framing, time pressure, and brand familiarity on download intentions for mobile apps, while also gathering qualitative data about the users' perceptions about the risk display.

Risk taking in the current study was operationalized as downloading apps that have high (low) risk (safety) scores, although in the real world, choosing to download an app that contains compromising permissions would indicate greater risk-taking behavior. The dependent variable in the experiment was whether an app, of the six shown on a given trial, was selected (binary: yes, no). This dependent variable was also justified due to its use in previous literature on security systems in mobile app stores (Chen et al., 2015; Rajivan and Camp, 2016; Chong et al., 2018). By using choice of app as the dependent variable, the predictors (security framing, time pressure, brand familiarity, and risk/safety score) were regressed onto a model by which we examined the most important factors in app selection.

CHAPTER 2

PILOT STUDY ON FAMILIARITY SCORES

The pilot study accomplished two goals. First, because there is no literature on the effect of brand familiarity on mobile app downloads, the familiarity ratings generated by the pilot study were used in the main experiment and can be used in future studies concerning app familiarity. Second, the correlations between perceived familiarity, favorability, and trustworthiness were examined in order to better understand the potential overlap between these three constructs in the mind of the mobile user.

2.1 Method

2.1.1 Participants. A total of 341 undergraduate participants from Old Dominion University's (ODU) psychology courses were recruited for the pilot study, 287 of whom completed the study. Participants were granted partial course credit or extra credit for their participation. The pilot was distributed in two versions, one containing 300 apps (25 types of apps x 12 apps per function; see Appendix A) and another that presented 150 apps, randomly displaying six of twelve total apps per function. These two versions of the study were used to avoid potential data-quality issue caused by the large number of apps in the 300-app version. The former version was completed by 190 participants (139 females, age *M* = 21.56, *SD* = 4.72) out of 220 total (86.36% completion rate) while the latter completed by 97 participants (78 females, one person declined to identify; age *M* = 22.03, *SD* = 5.12) out of 102 total (95.10% completion rate). The pilot study was granted exemption status from the IRB at ODU due to its benign methodology.

2.1.2 Materials. The study was hosted and accessed via Qualtrics (odu.qualtrics.com). The pilot study was comprised of 303 trials or 153 trials, depending on the pilot version (twelve or six apps per function listed in Appendix A), with three apps being attention checks. For the longer version of the pilot, all 303 apps were presented in a random order, while the shorter version was presented in 25 blocks with each block containing apps of a unique app function; within each block, six out of the twelve total apps were randomly selected and presented in random order. The second version of the pilot study was conducted to ensure that data quality did not decline due to the length of the study.

In the main questionnaires for both versions of the pilot study, each trial contained one app and three statements (“This app is familiar”, “I view the app favorably (I like this app)”, and “This app is trustworthy”), each on a seven-point Likert scale, with a score of one representing “Strongly Disagree”, seven representing “Strongly agree”, and four being “Neutral”. The three attention-check trials also contained apps, as participants would see for experimental trials, but contained a prompt for participants to rate the familiarity, the favorability, and the trustworthiness as “Disagree” to ensure they were paying attention. These attention-check trials were randomly presented amongst the legitimate trials. The pilot ended with an exit survey (see Appendix B) that included a multiple-choice question regarding factors that influence participants’ choice of apps (user ratings, icon look and feel, familiarity with the app or developer, and other), a question regarding the operating system of the participants’ personal cell phones (Android, iOS, other), a colorblindness check, and a demographic questionnaire (see Appendix C). The demographic questionnaire contained fields such as age, gender, race, color-blindness, and vision (normal or corrected-to-normal).

2.1.3 Procedure. At the beginning of the study, participants were instructed that they were to rate the familiarity, the favorability, and the trustworthiness of the apps. Afterwards, the participants rated each of the 303 or 153 apps (version dependent) on a seven-point Likert scale. After rating the apps, participants completed the aforementioned post-pilot survey. The longer version of the pilot study was completed in an average time of 84.31 minutes (*SD* = 253.21), while the shorter version took an average time of 37.64 minutes (*SD* = 76.41) to complete (note that these times are subject to error due to participants taking breaks during the study, distractions, etc., as two participants in the longer version each took over 45 hours to complete the pilot). Excluding just these two participants, the average amount of time for participants to complete the longer version of the pilot was 57.98 minutes (*SD* = 47.17), much closer to the median time of 41.05 minutes. After completing the pilot study, participants were granted credit for their participation.

2.2 Results

Of the 287 participants that completed the pilot study, 171 successfully answered at least two of the three catch trials. Due to the length of the study, we expected a degradation of vigilance, and thus worse performance at correctly identifying catch trials presented closer to the end of the study. The three catch trials were presented in random order amongst the other experimental trials and the relative catch trial presentation order was recorded (presented first, presented second, presented third). We therefore conducted a two-way ANOVA using randomization order (first, second, third) and pilot version (long, short) as predictors of correctly identifying catch trials for all 287 participants (see Figure 3). Interestingly, the participants' ability to correctly answer the catch trials did not degrade with time but improve, with the main effect of order being significant, $F(2,855) = 8.42$, $p < .001$, $\eta_p^2 = .02$. Pairwise comparisons revealed that participants were much more likely to miss the first catch trial (*M* = .50 proportion correct) than either the second (*M* = .61 proportion correct; *p* = .011) or third (*M* = .67 proportion correct; *p* < .001), while there were no significant differences between the second or third catch trials (*p* = .130). Meanwhile, the main effect of pilot version was not significant, *F*(1,855) = 2.32, *p* = .128, η_p^2 = .003, meaning that there was no statistical difference between the attentiveness of participants in the longer version of the pilot (*M* = .58 proportion correct) and those in the longer version of the pilot (*M* = .63 proportion correct). Furthermore, the interaction between catch trial order and pilot version was not significant, *F*(2,855) < 1, *p* = .984, η_p^2 = .000.

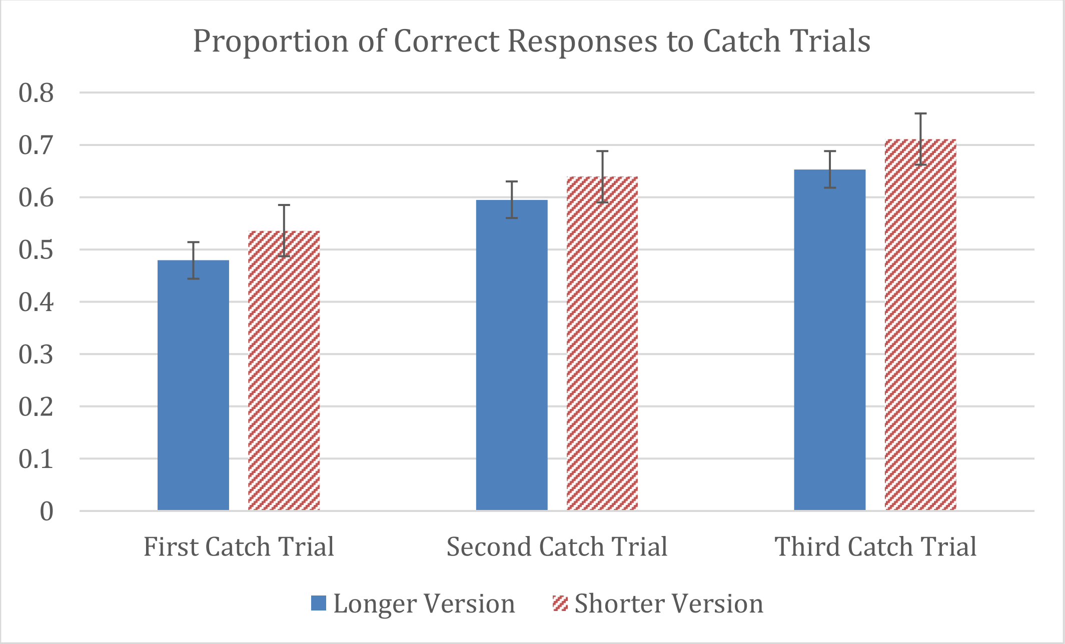


Fig. 3. Participants’ performance on catch trials as a function of trial order and pilot version. Error bars represent standard error of the mean.

In order to create brand familiarity scores, all data from both the longer and shorter versions of the pilot were combined. The Likert scale ratings (one through seven) provided by the 171 attentive participants were averaged for each app. The apps were then separated into their respective function categories and sorted from lowest brand familiarity rating to highest

(Appendix A). With the exception of the drawing app (which would become the catch trial in the main experiment), the apps with the three lowest scores and the apps with the three highest scores were chosen from each function for use in the main experiment. For the drawing category, the apps with the five lowest scores and the one highest score were chosen to serve as the catch trial for the main experiment. Collapsed across all app functions, the apps with the three lowest brand familiarity scores had an average score of 1.77 (*SD* = 0.30), averaged across app function. Conversely, the apps with the three highest brand familiarity scores had an average familiarity rating of 4.83 (*SD* = 1.25). The differences between the three lowest and three highest scores for all apps were examined using a one-way ANOVA with ranking (low, high) predicting brand familiarity scores. The results indicated a significant difference between the two groups, $F(1,148) = 423.88$, $p < .001$, $\eta_p^2 = .74$. By choosing apps with relatively low and high brand familiarity scores, respectively, the effect of brand familiarity on choice in the main experiment was expected to be maximized.

Correlations between familiarity, favorability, and trustworthiness were run for the apps. The resulting correlations were strong between all three constructs, from familiarity and favorability, $r(297) = .97$, $p < .001$, to favorability and trustworthiness, $r(297) = .99$, $p < .001$, to familiarity and trustworthiness, $r(297) = .96$, $p < .001$. These strong correlations indicate a high degree of overlap in participants’ ratings of the constructs.

We also analyzed the responses to the multiple-choice question on participants’ rationale when downloading apps (“user ratings”, “icon look and feel”, “familiarity with the app or developer”, and “other”; Chen et al., 2015) to see why the 287 participants choose apps on their own phones. Because participants were allowed to choose multiple reasons, Cochran’s Q test (1950) was conducted to examine the differences in how often a given reason was used. There was a significant difference in rationale, $\chi^2(3) = 295.02$, $p < .001$, with participants relying most heavily on “familiarity with the app or developer” (79.79% of participants), followed by “user ratings” (53.31% of participants), “icon look and feel” (43.55% of participants), and “other” (5% of participants. A pairwise post-hoc Dunn test with Bonferroni corrections ($\alpha = .013$) revealed significant differences between the frequency of each reason (*p*s < .001), except for the difference between icon look and feel and user ratings ($p = .028$).

2.3 Discussion

The pilot study was instrumental in laying a groundwork for research on consumer perception of the familiarity, favorability, and trustworthiness of mobile apps. By examining the correlations between these constructs, we have shown that there is merit in further exploring how they overlap and interact in forming opinions about apps. Granted, the methodology used in the pilot study still needs to be tested to ensure the validity and reliability of the correlative findings regarding familiarity, favorability, and trustworthiness of apps. Furthermore, the pilot indicated that mobile users are very likely to rely on the familiarity of an app when downloading from an app store, followed by the ratings provided by other users. In addition to these quantitative findings, the participants that chose the “other” category for app download rationale indicated that function and usefulness of the app is important in their decision-making process (see Appendix D for participants’ responses for the “other” category). Participants indicated that they were more likely to download an app if it had been marketed to them through advertisements or recommended by friends or family.

The significant difference between generated familiarity scores for the expected high and low familiarity groups confirmed the app selection by the experimenters for the following experiment. However, the pilot study showed that certain app functions with heavier expected use (e.g., social media, web music streaming, etc.) showed a much greater discrepancy between the expected high and low familiarity apps, while the differences in familiarity were not as strong for apps of more niche audiences (e.g., fitness, drawing, etc.). Most importantly, the brand familiarity ratings generated by the pilot study can serve as a new scale by which researchers can measure this construct, as was done in the following experiment. As previously discussed, these familiarity scores are a novel measure and, given that there is no previous literature on brand familiarity for mobile apps, the theoretical and practical implications for such measures are ripe with opportunity.

CHAPTER 3

MAIN EXPERIMENT

This experiment aimed to further investigate the security score system proposed by Chen et al. (2015), the role that time pressure plays in the decision-making process, and the degree to which brand familiarity predicts selection preference. All participants were presented with multiple choice screens that contain an assortment of apps. Participants were assigned to one of four conditions according to the framing of security scores of the apps (safety, risk) and time pressure assignment (present, absent). Security scores were framed as either safety using closed locks or risk using open locks, consisting of one, two, three, four, or five locks; safety scores indicated higher levels of safety (less risk) with increasing locks, while risk scores indicated higher levels of risk (less safety) with increasing locks. As discussed in the pilot study above, the brand familiarity scores were determined before the experiment by the pilot study. Whereas the previous study by Chen et al. (2015) excluded the most recognizable apps, the current study used both familiar and unfamiliar apps, each with an assigned value of familiarity for analyses. Furthermore, this study was the first to introduce time pressure to the literature on mobile app downloads. The dependent variable was whether an app is chosen (yes, no; 1, 0).

Hypothesis 1.1: For the main effect of security scores on choice, increases in safety scores (gain frame) would positively predict app choice, such that, with each additional lock, a participant would be more likely to download an app; conversely, increases in risk scores (loss frame) would negatively predict choice, such that, with each additional lock, the app would be less likely to be chosen (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018).

Hypothesis 1.2: For the interaction between security framing and security score on choice, the security scores under the safety frame would be more impactful on choice (i.e., safer apps would be chosen more and less safe apps would be chosen less), while security scores under the risk frame were expected to less clearly guide decision making, possibly due to a confusion of the score's meaning (Chen et al., 2015). Using the terms of the framing effect, the anticipated response was more risk aversion when using a safety/gain frame than when using a risk/loss frame (Tversky & Kahneman, 1979).

Hypothesis 2a: For the three-way interaction between time pressure, security framing, and security scores on choice, per the research by Saqib and Chan (2015), there would be a reversal in the framing effect under time pressure. Without time pressure, the security scores under the safety framing would be more effective (i.e., lead to risk aversion) at guiding choices than the security scores under the risk framing (see Hypothesis 1). However, with time pressure, the security scores under the risk frame would more effectively impact choice, with participants choosing apps of stronger security score than those in the safety frame.

Hypothesis 2b: Alternatively, for the interaction between time pressure, security framing, and security score on choice, instead of a reversal in risk preference, the existing discrepancy between the effectiveness of security scores under safety and risk frames could be exaggerated under time pressure. In Chen et al.'s (2015) study, participants failed to fully understand how the risk scores worked. If participants in the current study were also confused by their presentation without time pressure, the differences between the two security score frames were expected to be magnified under time pressure. In other words, the risk framed security scores may have resulted in confused responses, while the safety framed security score's simplicity may have been more conducive to rapid-fire decision making; the participants under time pressure would choose apps with higher safety scores more frequently than other apps, whereas the risk scores would provide participants with very little information under time pressure, rendering the security system ineffective.

Hypothesis 3: For the main effect of time pressure on decision times, decision times would be faster for those in the time pressure condition than those not under time pressure (Madan et al., 2015).

Hypothesis 4: For the main effect of brand familiarity on choice, brand familiarity would positively predict choice, with less familiar apps having a lower likelihood of download, due to a greater sense of trust born out of familiarity (Ha & Perks, 2005).

Hypothesis 5: For the interaction between brand familiarity and time pressure, compared to participants without time pressure, the association between brand familiarity and likelihood of download would be stronger for those under time pressure. This result was expected due to browsing behaviors shown by Liu et al. (2017), wherein participants under time pressure focused more on familiar brands with greater observation durations and counts than on their competitors.

Hypothesis 6: There would be an interaction between security scores and brand familiarity. Based on the strong effect of brand familiarity (Ha & Perks, 2005), it was expected that the effect of security score on likelihood of download would be stronger for apps with lower brand familiarity than for those with higher brand familiarity. That is to say, participants would be more reliant upon safety (risk) scores for apps with lower brand familiarity ratings than the more familiar apps, if they decided to download apps with low brand familiarity scores. Participants were expected to be more risk taking with familiar apps, not relying on the security scores. However, the lack of literature on this interaction means that this hypothesis was exploratory.

3.1 Method

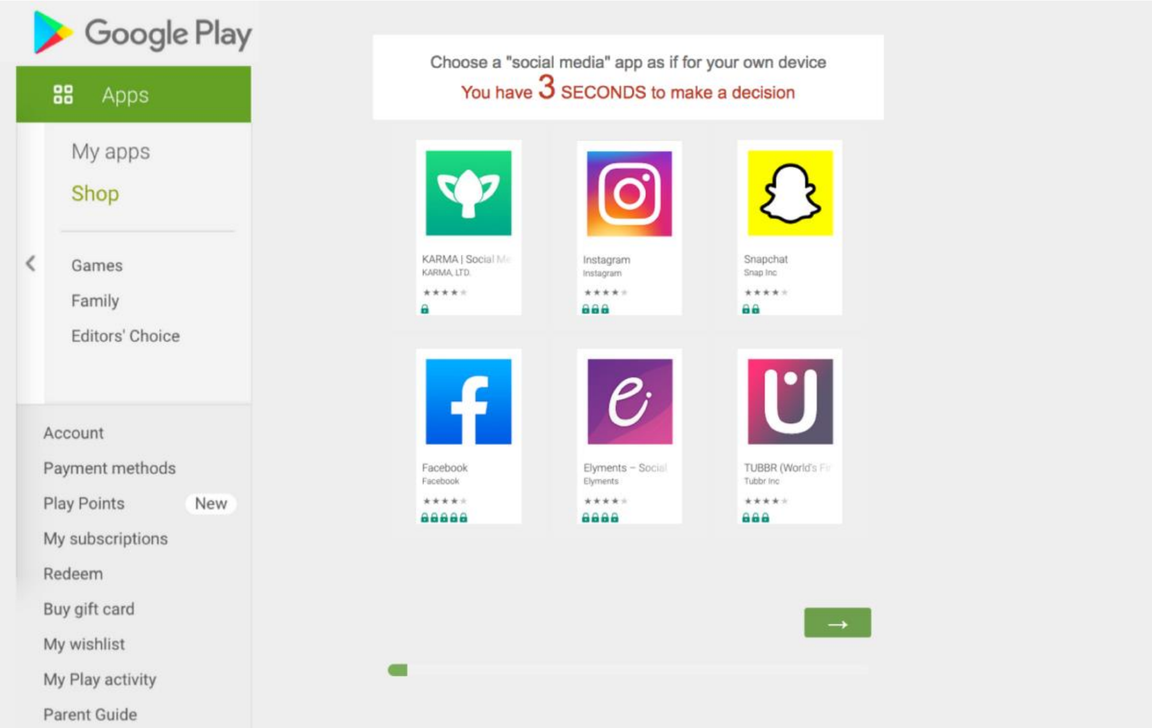
3.1.1 Participants. Based on the effect size (η^2 = .01) provided by previous research (Chen et al., 2015), a power analysis for a generalized linear regression was conducted in G*Power with power set at .80 (see Appendix E for power analysis). However, because of the design of the experiment (its use of nested levels), the variance both between and within participants was considered under the framework of hierarchical linear modeling (HLM; Lindley & Smith, 1972). Therefore, an Intraclass Correlation (ICC) was calculated using the current study's data to examine such variance. The ICC was then used in West, Ryu, Kwok, and Cham's (2011) formula, along with the total number of apps participants would see, to determine the number of participants required to attain the necessary independent data points (the output created by G*Power). Due to the small value of the ICC (.021) from the current study, the total number of legitimate apps seen (144), and the output by G*Power (124 independent data points), the effective number of individual data points (143.89) would be captured in four participants' results (See Appendix F for analysis). Therefore, there was little concern for strong individual differences that would skew the results of the current study.

Participants were recruited via Amazon's Mechanical Turk (MTurk) and were compensated one dollar each for their participation. MTurk is an online platform wherein the general population can participate in online studies for compensation. The practice of recruiting participants through MTurk has been shown to produce quality results, especially when experimenters restrict the qualifying parameters to only include the most attentive of participants (Peer, Vosgerau, & Acquisti, 2014). We thus recruited 128 participants (51 females; age *M* = 40.96, *SD* = 12.25) for the main experiment, though demographic data from two of the participants were incomplete, resulting in 126 participants for the analyses on choice and 128 for the analyses on decision time. Demographic data were incorporated into the choice analyses in order to ensure there were no confounds but were not included in decision time analyses because there was no expectation of confounds. Thus, the incomplete demographic data for these two participants resulted in their exclusion in the choice analyses. Participants in the non-time pressure condition completed the experiment in an average time of 11.86 minutes (*SD* = 5.16), while participants in the time pressure condition completed the experiment in an average time of 9.75 minutes (*SD* = 5.60). MTurk participants were required to live in the United States of America and to have a 95% HIT approval rate in order to ensure quality data collection. Due to the benign nature of the study's methodology, it was granted exemption status from the IRB at ODU.

3.1.2 Design of experiment. The independent variables included brand familiarity scores (as determined by the pilot study), security framing (risk, safety), security score (one, two, three, four, or five locks), and time pressure (present, absent). Security framing and time pressure were between-subjects, while security scores and brand familiarity scores were within-subjects. Security scores for apps were assigned a safety (risk) score from one to five, with all five types of scores present during each trial and the middle scores (three locks) represented twice each. The apps within each trial contained familiarity ratings, invisible to the participants, according to the pilot study, that acted as a predictor variable when analyzed. Finally, user ratings were controlled by assigning a rating of four stars across all apps to minimize their effect on app choices. Participants were provided 24 experimental trials (Schuster et al., 2015) and one catch trial. The catch trial consisted of five apps with low brand familiarity and low (high) safety (risk) scores and one app with high brand familiarity and a high (low) safety (risk) score. In this way, only inattentive participants were expected to choose any of the five unattractive apps. Because risk taking behaviors were the focus of this study, the dependent variables were decision times and whether or not each app was chosen, given its security score and the familiarity of the brand. Choices on all six apps were recorded, five of which had a dependent variable value of zero (not chosen), while one had a score of one (chosen). The decision time on a trial was recorded from the beginning of the trial until participants clicked to advance the page. First click responses were not analyzed since participants could change their choices. Last click responses were not analyzed due to technical errors within Qualtrics wherein some last click responses were not logged.

3.1.3 Apparatus. The experiment took place online. Participants used their own devices (laptops or desktops) to ensure the proper display of the stimuli in the experiment and similar engagement across participants. Participants were required to use traditional computers as opposed to mobile devices, such as phones or tablets. Devices were controlled by Qualtrics, such that mobile users were excluded from the experiment before it began. By equating the types of devices used for the experiment, all participants saw the prompts in the same manner and their subsequent responses were not be altered by any orientation artifacts.

3.1.4 Materials. The study was hosted and accessed via Qualtrics (odu.qualtrics.com). The experiment was comprised of the same demographic questionnaire as from the pilot study (Appendix C), 25 app download screens (one was a catch trial), and an exit survey. For the app download screens, the design replicated the current desktop version of the Google Play Store (see Figure 4). The function of the apps varied from trial to trial, from social media to note-taking to weather apps, but each trial was uniform in function (Appendix A). The desktop version was used instead of the mobile version in order to better fit the orientation of desktop and laptop devices used, as well as to better simulate the environment a user would see using such a device.



Choose a "social media" app as if for your own device

You have 3 SECONDS to make a decision

Fig. 4. Example of app store in Experiment 1 using the safety score system. Six apps were randomly sorted in each trial, with each trial consisting of apps with similar functions (e.g., social media apps). Each app had a user rating (controlled at four stars for all apps) and a security score (randomized across apps), framed as risk (negative) or safety (gain, shown). Finally, participants in the time pressure condition saw a countdown timer, as shown, above the apps. Those in the non-time pressure condition saw the same interface, excluding the red countdown timer below the app function description at the top of the screen.

Brand familiarity ratings were assigned to the respective logos for each of the six apps based on data from pilot, while app location was randomized across the six spaces to prevent any confounds caused by ordering effects, wherein the first visible option could be more attractive than others or wherein a static location for the highest brand familiarity score would anchor subsequent responses. Security scores were assigned to the apps in ascending order and utilized a Latin Square design, with six different versions of the trial. Participants were randomly shown one of the six trial versions for each app function. For example, if apps "A" through "F" were static for coding purposes, version one of the trial would assign one lock to "A", two locks to "B", and so on until "F" had five locks. Version two would assign one lock to app "B", four locks to app "F", and five locks to app "A"; this process would continue for all six versions. This allowed all apps to have all possible security scores.

The security score system was framed as riskiness or safety and consisted of colored locks (see Figure 5 for the security score system), with risk and safety being inversions of each other (e.g., a risk score of two would be a safety score of four). Those in the safety frame saw a scale from one to five teal locks (with higher scores meaning safer apps) and those in the risk frame saw a scale from one to five pink locks (with higher scores meaning greater riskiness). Teal and pink coloring were chosen because of their similarity to green and red, respectively; green and red have associated "go and stop" meanings (Bergum & Bergum, 1981), but can be nearly indiscernible for those with red-green color blindness.

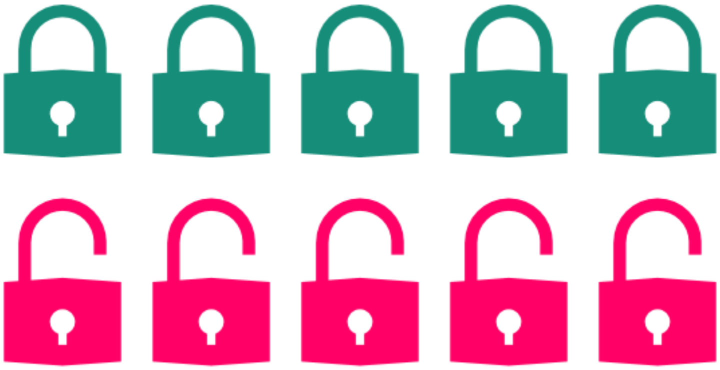


Fig. 5. Security framing used teal and pink coding for increased salience. These colors are similar to green and red and may still be reminiscent of stop and go signaling (Bergum & Bergum, 1981), while considering those with color blindness. Locks were chosen for dual coding

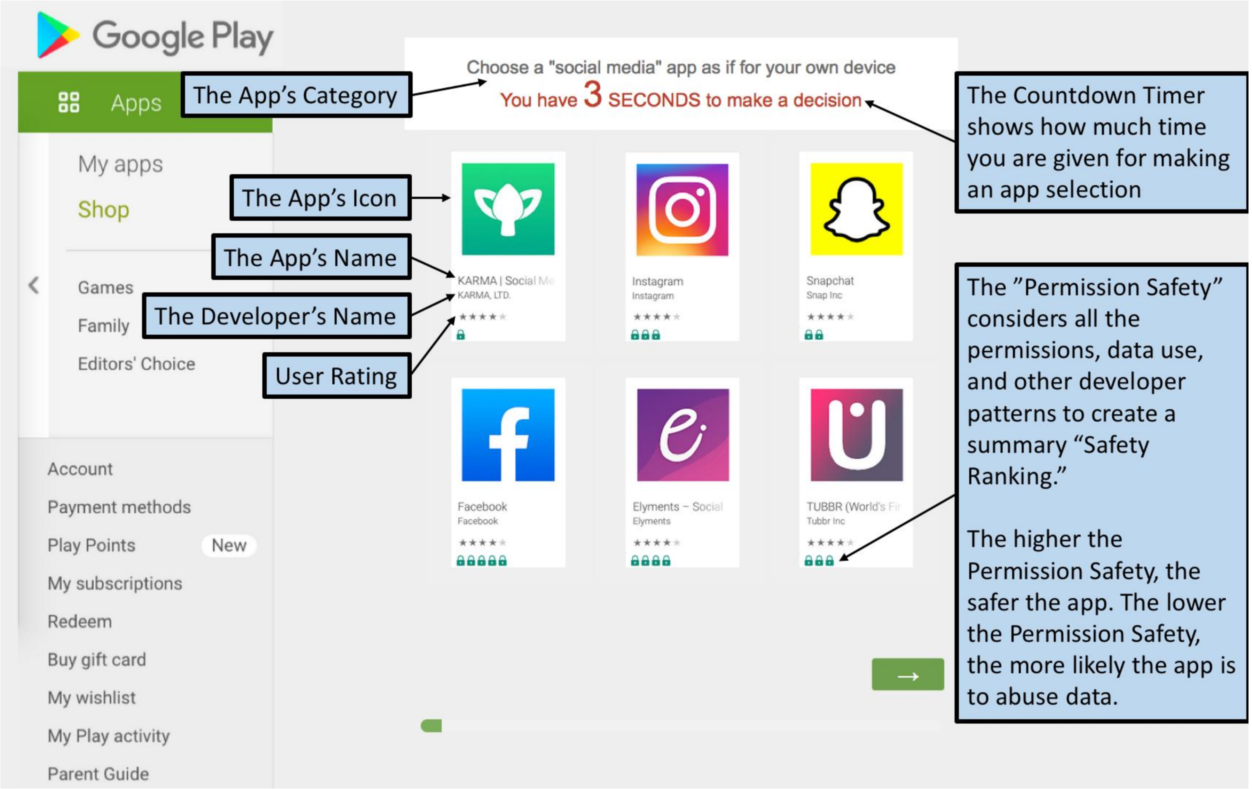
purposes, with the locked locks indicating safety and unlocked locks suggesting risk (Rajivan & Camp, 2016).

Participants assigned to the time pressure conditions saw countdown timers within the Google Play Store interface. The time for each countdown timer was determined by collecting the means for each trial generated by those in the non-time pressure conditions. The time allotted for each trial's countdown timer were equal to the mean decision time; mean time was chosen instead of the mean minus the standard deviation because of the large size of the standard deviation values. The use of mean times is supported by previous literature as well (Chen & Proctor, 2017). Finally, the exit survey examined participants' subjective reasoning for download decisions and thoughts regarding the experiment (see Appendix G for the survey).

3.1.5 Procedure. The first half of the participants were assigned to the non-time pressure condition with randomly assigned security score type (safety score, risk score). The mean decision times provided by the non-time pressure participants were then used for the time pressure participants. Because of this need for countdown times, the time pressure participants were recruited afterwards and randomly assigned security score type. While those in the non-time pressure condition took as much as they needed to respond to the questions, those in the time pressure condition were shown countdown timers to encourage expediency. If participants took longer to make decisions than the time provided by the timer, the timer ended and showed feedback asking them to respond faster. This setting allowed for full data collection and was expected to still induce time pressure despite the lack of consequence for slow responses.

Participants were instructed to choose an app per trial as if selecting an app for their personal mobile device; see Figure 6 for instruction screen example. Participants had to make decisions based solely on the available information provided on the screen of the device.

Participants were able to click the app of their choosing, advancing to the next screen after each selection.



Choose a "social media" app as if for your own device

You have 3 SECONDS to make a decision

The Countdown Timer shows how much time you are given for making an app selection

The "Permission Safety" considers all the permissions, data use, and other developer patterns to create a summary "Safety Ranking."

The higher the Permission Safety, the safer the app. The lower the Permission Safety, the more likely the app is to abuse data.

Fig. 6. Example of the instruction page participants in the time pressure condition saw before the experiment begins. Those without time pressure saw the same image without the red countdown timer below the App Category prompt. Textboxes contained descriptions of each component of the Google Play Store.

After completing all 25 trials, participants were administered a questionnaire to collect subjective reasoning for their choices and any thoughts they had about the experiment. The questionnaire consisted of rationale questions for app selection, the CyberDOSPERT (Kharlamov, Jaiswal, Parry, & Pogrebna, 2018), an attention check that asked the meaning of the locks, a colorblindness check, Likert scale questions regarding cybersecurity expertise, an open-ended prompt for additional feedback. The CyberDOSPERT (Kharlamov et al., 2018) is a domain-specific scale designed to measure risk tolerance for cyber practices (sharing passwords, using public WiFi, etc.); if participants scored high on this scale, it may influence individual risk-taking behaviors. Upon completion of the questionnaire, the study was finished and participants were compensated.

3.2 Results

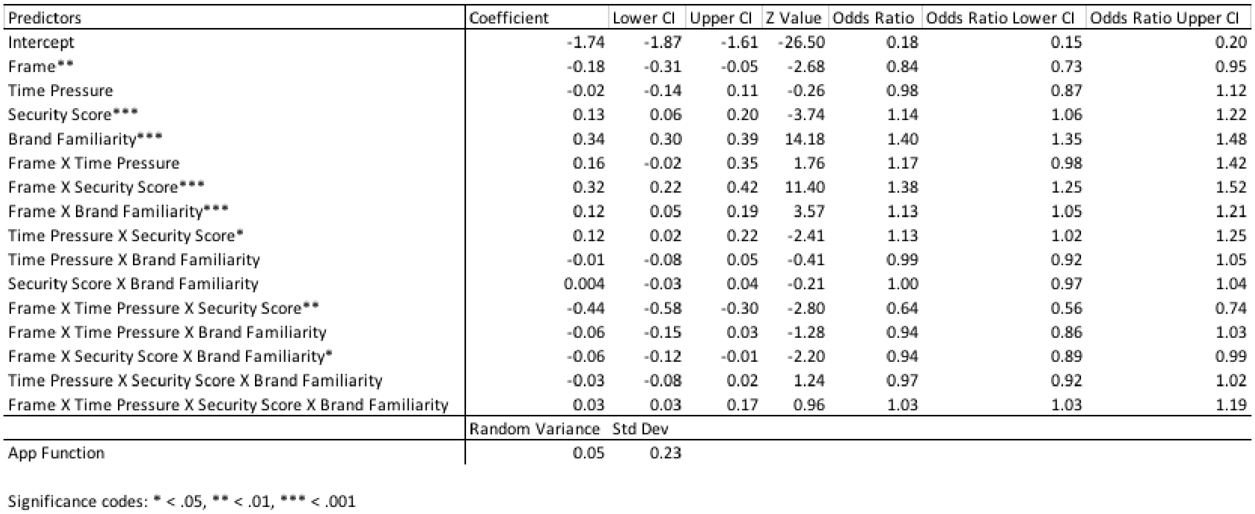
3.2.1 Choice Data. Based on the ICC calculated for the current study's data, a higher-order analysis, such as HLM (Lindley & Smith, 1972) was deemed unnecessary. Rather, a generalized linear mixed-effects regression with the random intercept effects for participants and app function was used in R to analyze choice data (Baayen, 2008; Chen et al., 2018). Participants and app function served as random factors to account for variability not explained by brand familiarity, security score, time pressure, or security framing. By analyzing these variables with a generalized linear mixed-effects regression, we were able to determine the degree to which each variable impacted app selection. The random factor variance (similar to ICCs in HLM) for participants was zero, so the analysis was conducted using only app function as the random factor on app choice. Security scores (Hypothesis 1.1) and brand familiarity scores (Hypothesis 4) served as within-subject predictor variables of app choice, while time pressure and security framing were between-subject predictor variables, with the criterion variable being app selection (coded 0 or 1). Both security scores and brand familiarity scores were treated as continuous variables and centered before inclusion in the model. Interaction terms were also included within the R script for the hypotheses for the security scores X framing interaction (Hypothesis 1.2), the security scores X framing X time pressure interaction (Hypotheses 2a/2b), the time pressure X familiarity interaction (Hypothesis 5), and the brand familiarity X security scores interaction (Hypothesis 6).

The catch trial was designed to flag participants that did not pay attention by including five low brand-familiarity apps with one-star safety (five-star risk) ratings and one high brand-familiarity app with a five-star safety (one-star risk) rating. Unfortunately, due to a sizeable number of participants failing to choose the familiar, safe app during the catch trial (110 out of 128), the catch trial was deemed uninformative due to participants being unable to distinguish the catch trial and consequently all participants' data were included. Two participants' choice data were excluded from analysis due to incomplete data acquisition in the demographics portion of the experiment, as stated previously in **Participants** (1.6% of total participant data). Additionally, the trial containing the weather apps failed to collect choice data for 20 participants; all other data for these 20 participants were included in the analysis. Out of the 126 viable participants, the uncollected weather app data from 20 participants accounted for 0.66% of the decision data. Data for the remaining participants' choices for the weather apps were retained for analysis.

A generalized linear mixed-effects regression (GLMER) with the random intercept effect for app function revealed the predictive slopes of each predictor of choice (see Table 2). Note that the coefficients in the table are in relation to the reference group (participants in the risk frame not under time pressure) and that coefficients reported are therefore differences between the predictor or interaction of predictors from the intercept of -1.74 in Table 2. Demographic information, such as age, gender, ethnicity, CyberDOSPERT scores, personal mobile device, self-reported cybersecurity expertise, and colorblindness, were included in the model as predictor variables; these variables and all possible interactions with other predictor variables were created within the analysis to ensure there were no confounding variables that predicted choice. None of the demographic information significantly predicted choice, *p*s >.05.

Table 2.

Coefficients in the General Linear Mixed Effects Regression



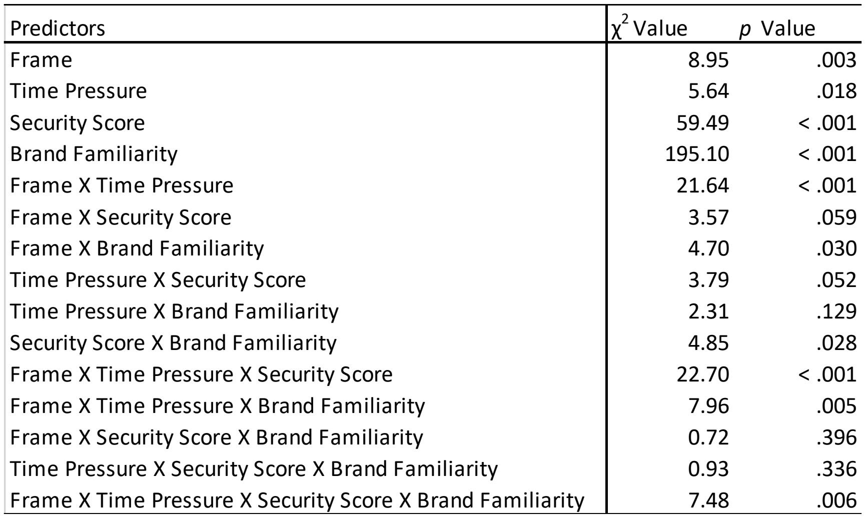
Predictors	Coefficient	Lower CI	Upper CI	Z Value	Odds Ratio	Odds Ratio Lower CI	Odds Ratio Upper CI
Intercept	-1.74	-1.87	-1.61	-26.50	0.18	0.15	0.20
Frame**	-0.18	-0.31	-0.05	-2.68	0.84	0.73	0.95
Time Pressure	-0.02	-0.14	0.11	-0.26	0.98	0.87	1.12
Security Score***	0.13	0.06	0.20	-3.74	1.14	1.06	1.22
Brand Familiarity***	0.34	0.30	0.39	14.18	1.40	1.35	1.48
Frame X Time Pressure	0.16	-0.02	0.35	1.76	1.17	0.98	1.42
Frame X Security Score***	0.32	0.22	0.42	11.40	1.38	1.25	1.52
Frame X Brand Familiarity***	0.12	0.05	0.19	3.57	1.13	1.05	1.21
Time Pressure X Security Score*	0.12	0.02	0.22	-2.41	1.13	1.02	1.25
Time Pressure X Brand Familiarity	-0.01	-0.08	0.05	-0.41	0.99	0.92	1.05
Security Score X Brand Familiarity	0.004	-0.03	0.04	-0.21	1.00	0.97	1.04
Frame X Time Pressure X Security Score**	-0.44	-0.58	-0.30	-2.80	0.64	0.56	0.74
Frame X Time Pressure X Brand Familiarity	-0.06	-0.15	0.03	-1.28	0.94	0.86	1.03
Frame X Security Score X Brand Familiarity*	-0.06	-0.12	-0.01	-2.20	0.94	0.89	0.99
Time Pressure X Security Score X Brand Familiarity	-0.03	-0.08	0.02	1.24	0.97	0.92	1.02
Frame X Time Pressure X Security Score X Brand Familiarity	0.03	0.03	0.17	0.96	1.03	1.03	1.19
	Random Variance	Std Dev					
App Function	0.05	0.23					

Significance codes: * < .05, ** < .01, *** < .001

Also note that coefficients in Table 2 do not represent significance; likelihood ratio tests (LRT) were conducted for each of the model's terms to test the significance of the main effects and interactions of the various predictors (see Table 3). In order to determine if a term is significant, the LRT compares the original model to another model that omits the term in question using a Chi-Square test. For example, in order to test the significance of brand familiarity, the original model, which includes the term, is compared to a model wherein brand familiarity has been removed as a predictor. Therefore, while the GLMER's coefficients provide information regarding the differences in behavior between groups, the LRT provides information about the strengths of the predictors themselves rather than in relation to a specific group. The combination of both tests is critical to understanding the study's results.

Table 3.

Likelihood Ratio Test Results for GLMER Terms



Predictors	χ^2 Value	*p* Value
Frame	8.95	.003
Time Pressure	5.64	.018
Security Score	59.49	< .001
Brand Familiarity	195.10	< .001
Frame X Time Pressure	21.64	< .001
Frame X Security Score	3.57	.059
Frame X Brand Familiarity	4.70	.030
Time Pressure X Security Score	3.79	.052
Time Pressure X Brand Familiarity	2.31	.129
Security Score X Brand Familiarity	4.85	.028
Frame X Time Pressure X Security Score	22.70	< .001
Frame X Time Pressure X Brand Familiarity	7.96	.005
Frame X Security Score X Brand Familiarity	0.72	.396
Time Pressure X Security Score X Brand Familiarity	0.93	.336
Frame X Time Pressure X Security Score X Brand Familiarity	7.48	.006

Of importance to the study at hand, security scores were a significant predictor of choice, supporting Hypothesis 1.1, $\chi^2(1) = 59.49$, $p < .001$; the coefficient was 0.13, with a 95% confidence interval (CI) of [0.06, 0.20] (see Figure 7). When transformed from a log likelihood into an odds ratio, the results indicate that, with each additional (reduced) safety (risk) lock, users are 1.14 times more likely to download a given app, 1.14, CI [1.07, 1.22].

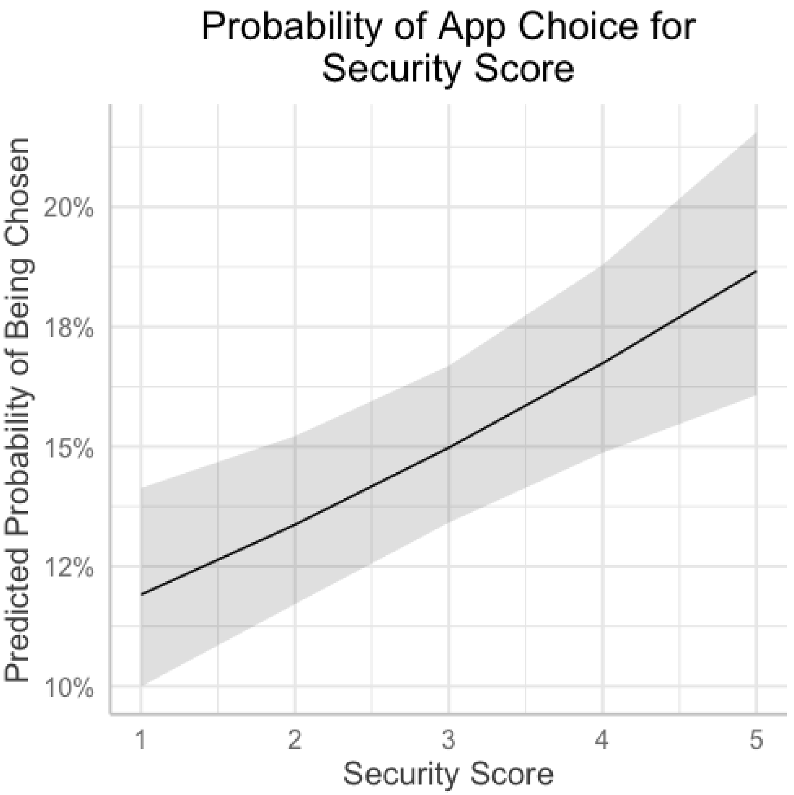


Fig. 7. The effect of security scores on app choice. With increasing safety (decreasing risk) scores, participants were significantly more likely to choose an app.

Conversely, the interaction between security scores and framing did not significantly predict choice, meaning that Hypothesis 1.2 was not supported, $\chi^2(1) = 3.57$, $p = .059$; the coefficient was 0.32 CI [0.22, 0.42] (see Figure 8). Despite the lack of significance, predicted probability of choice was calculated for security scores (one through five) by framing (safety, risk) and indicated a steeper change in probability of choice for the safety framing (.08, .11, .14, .18, .22) than the risk framing (.11, .13, .15, .17, .20) for security scores. Note that this finding is in the same trajectory as other studies, with safety framing having a stronger effect than risk framing on choice; one reason for the nonsignificant finding for the interaction between security scores and framing could be because of the time pressure conditions were averaged for the analysis.

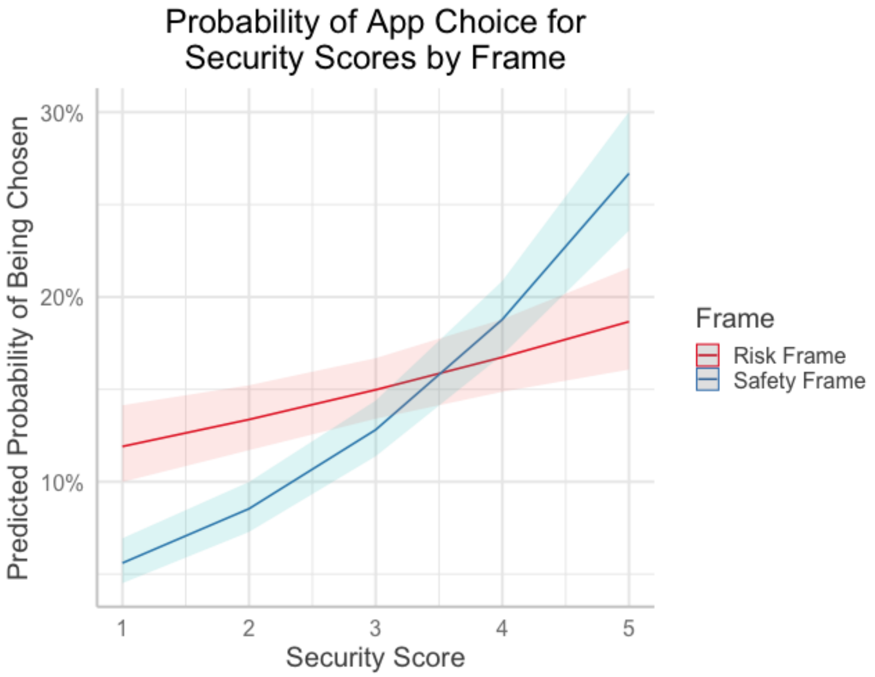


Fig. 8. The effect of the interaction between security scores and frame on app choice. While both increases in safety ratings and decreases in risk ratings predicted the selection of apps, the safety frame was significantly more effective at guiding choice than the risk frame.

Indeed, the interaction between time pressure, security framing, and security scores was a significant predictor of choice, $\chi^2(1) = 22.70$, $p < .001$; the coefficient was -0.44 CI [-0.58, -0.30]. When transformed from a log likelihood into an odds ratio, the results indicate that participants under time pressure were 1.2 (1/0.83 odds ratio) times less likely to choose the safety scores than participants without time pressure who experienced the risk scores with increasing (decreasing) locks, 0.83, CI [0.56, 0.74]. As was done for the two-way interaction, predicted probability of choice was calculated for security scores (one through five) by framing (safety, risk) and time pressure (absent, present). Without time pressure, the safety framing had a steeper slope (.06, .09, .13, .19, .27) than the risk framing (.12, .13, .15, .17, .19). However, with time pressure, the safety framing was less steep (.12, .13, .15, .16, .18) than the risk framing (.09, .12, .15, .18, .22). Therefore, while participants that were not under time pressure relied more heavily on safety ratings than risk ratings, the effectiveness of the safety framing was no better than the risk framing under time pressure. Indeed, the introduction of time pressure means that increases in safety scores were no more impactful than equivalent decreases in risk ratings. In fact, the difference in the effectiveness of safety rating between time pressure and no time pressure was so great that, under time pressure, the risk ratings led to safer decision-making, supporting Hypothesis 2a.

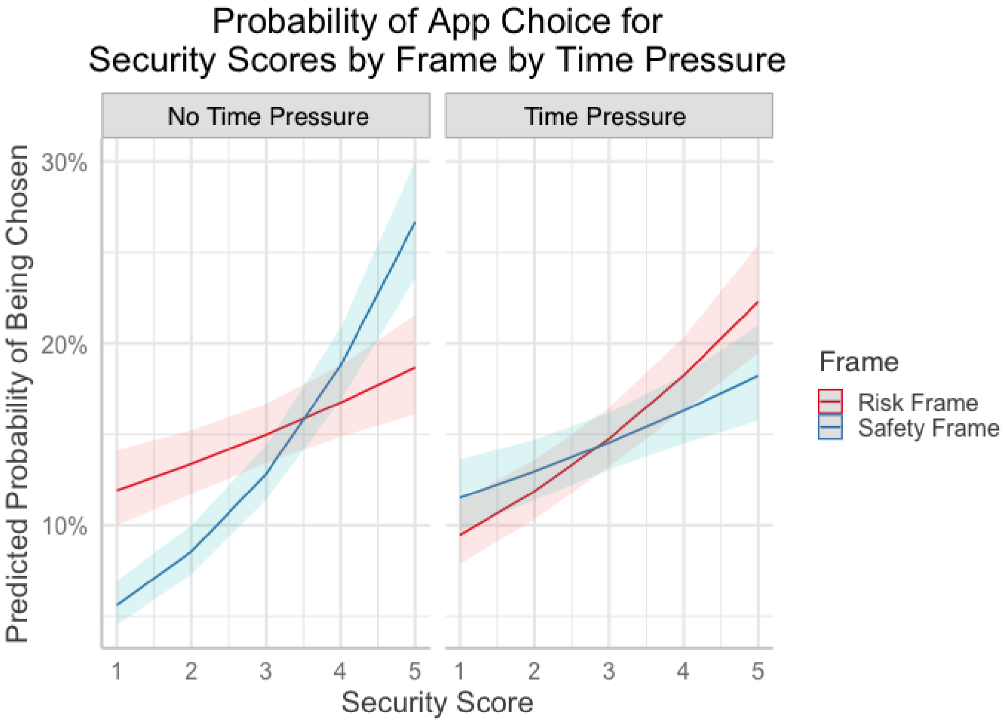


Fig. 9. The effect of the interaction between security scores, security frame, and time pressure on app choice. Without time pressure, the safety framing interacts with security scores in a stronger fashion than the risk framing, such that increases in safety ratings are much more effective at guiding choice than decreases in risk ratings. However, under time pressure, this advantage disappears, if not reverses.

Brand familiarity was a significant predictor of choice, $\chi^2(1) = 195.10$, $p < .001$; the coefficient was 0.34 CI [0.30, 0.39], with more familiar brands being chosen more often, supporting Hypothesis 4 (see Figure 10). Transformed into an odds ratio, this means that, with each increasing unit of brand familiarity score (out of seven total), users were 1.35 times more likely to download a more familiar app, 1.35, CI [1.35, 1.48].

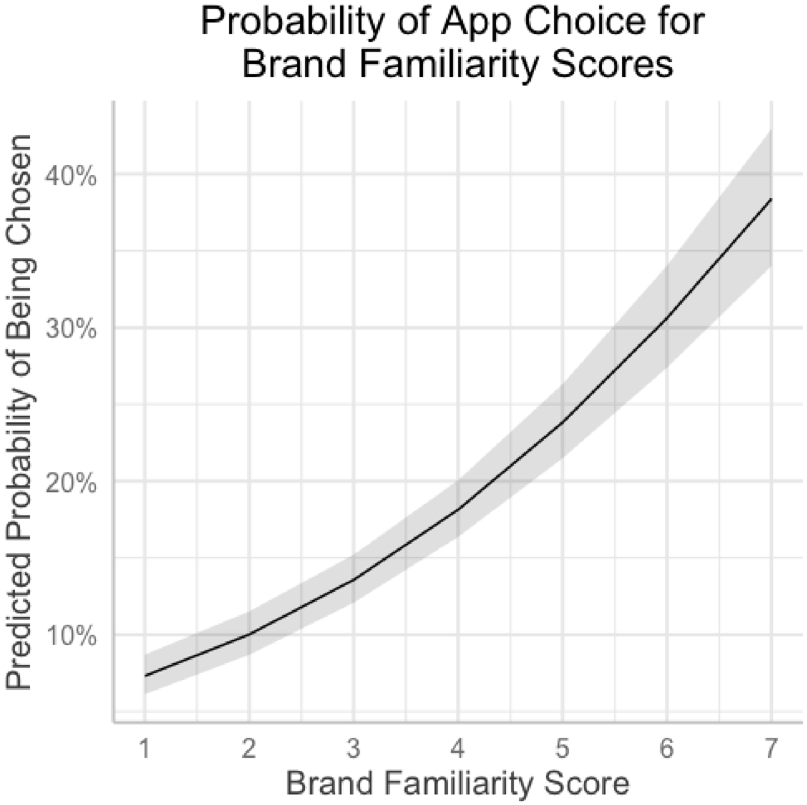


Fig. 10. The effect of brand familiarity on app choice. With increasing brand familiarity, per the scores created in the pilot study, participants were more likely to choose an app.

However, the interaction between brand familiarity and time pressure did not significantly predict choice, $\chi^2(1) = 2.31$, $p = .129$; the coefficient was -0.01 CI [-0.08, 0.05], meaning that Hypothesis 5 was not supported (see Figure 11). This may be due to a ceiling effect, wherein familiar brands were already chosen so often that there was little room for participants to choose familiar brands even more frequently under time pressure. While the probability that an app with a brand familiarity score close to seven was around 40%, keep in mind that this was within a model that contained several other significant predictors of choice.

That is to say, the ceiling effect of brand familiarity averages the effect of all other predictors, such that a brand familiarity score of seven, for example, contains the entire range of security frames and scores (safety and risk; one through five; see Figures 10 and 11). This averaging across all other predictors and the inclusion of their own respective predictive strengths in the model is why the ceiling for brand familiarity may peak around 40%.

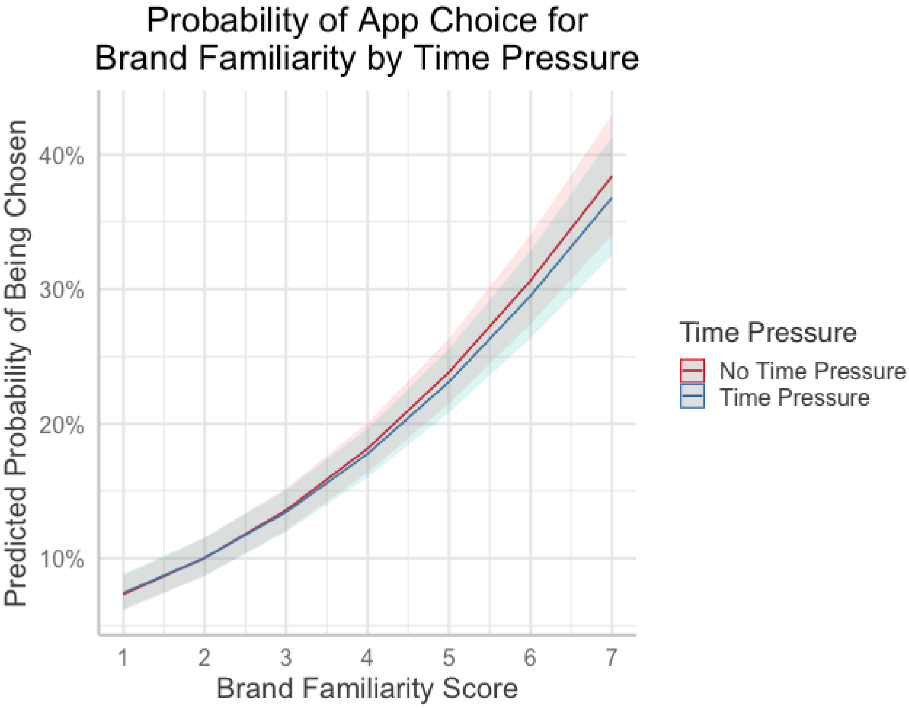


Fig. 11. The effect of the interaction between brand familiarity and time pressure on app choice. The interaction between brand familiarity and time pressure was not significant, with little difference between the non-time pressure condition and the time pressure condition. Indeed, the similarity between Figure 10 and Figure 9 indicate negligible impact of time pressure on the effect of brand familiarity on choice.

Finally, the interaction between security scores and brand familiarity significantly predicted choice, $\chi^2(1) = 4.85$, $p < .028$; the coefficient was 0.004 CI [-0.12, -0.01], though not in the direction that Hypothesis 6 predicted (see Figure 12). This significant finding indicates that, with increasing brand familiarity, participants changed the way in which they interacted with the security scores. This interaction appears to be due to a floor effect wherein participants were unlikely to download unfamiliar apps enough to fully utilize the security scores. Nevertheless, apps with higher brand familiarity were subject to more discrimination with regard to the security score. This is to say, rather than choosing low-familiarity apps, participants were more likely to default to more familiar apps and use security scores to determine which of the familiar apps they would download. Again, this increase in difference between security score effectiveness as brand familiarity increased could be due to the low rate at which low-familiarity apps were chosen, but suggests that security scores could, in fact, guide consumer purchases with familiar apps.

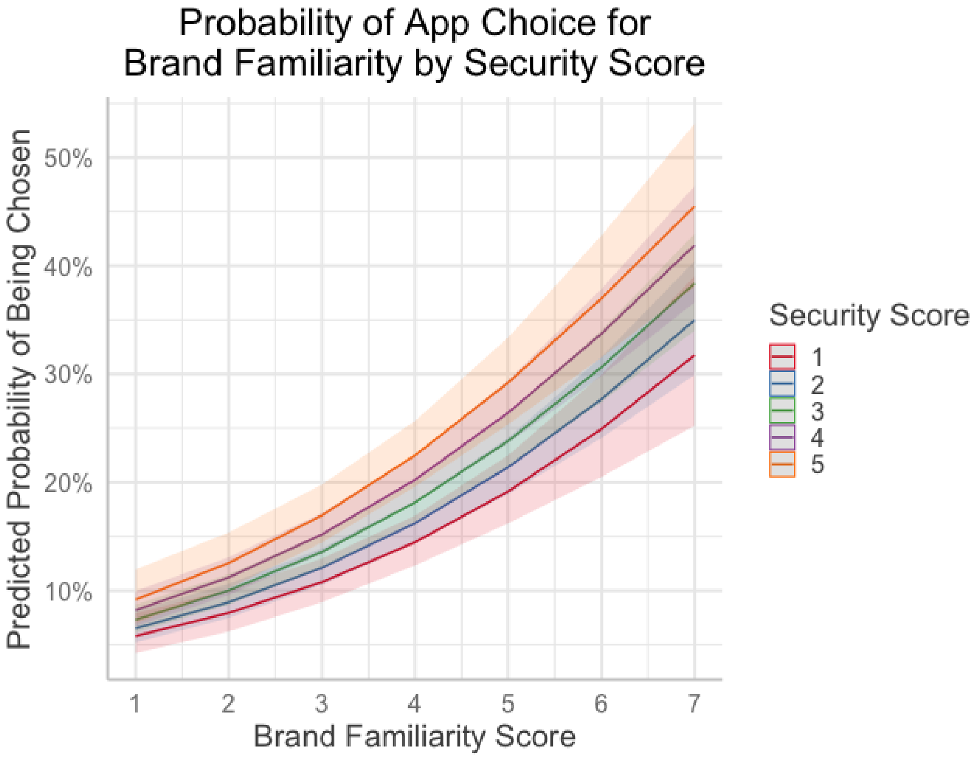


Fig. 12. The effect of the interaction between security scores and brand familiarity on app choice.

In contrast with our expectations, participants' willingness to choose riskier apps did not increase with higher brand familiarity. As can be seen, participants still chose apps with higher (lower) safety (risk) scores despite the level of brand familiarity.

3.2.2 Decision Time Data. For decision time, a two-way ANOVA was run, with time pressure (present, absent) and security framing (risk, safety) as independent variables and decision time as the dependent variable. Distribution of decision times was not normal, but skewed right and leptokurtic. This means that participants generally made decisions quickly, though there were a few very slow decision times that made the sample heteroskedastic. To account for this lack of normality and extremely slow times, decision time data were Winsorized at the fifth and ninety-fifth percentiles; that is, data points outside these percentiles were transformed to be equal to these percentiles, reducing the number of extreme decision times. A total of 6.25% of the decision time data were Winsorized to fit within the defined range. Decision times were then log-transformed (natural log) for analysis (Chen & Proctor, 2017); note that the reported means and standard deviations are Winsorized decision times (in seconds) rather than log-transformed values for the purpose of easier understanding. The two-way ANOVA revealed a significant main effect of time pressure, $F(1,3196) = 487.32$, $p < .001$, $\eta_p^2 = .13$, with participants under time pressure making faster decisions ($M = 4.87$ s, $SD = 3.07$ s) than those without time pressure ($M = 8.35$ s, $SD = 7.06$ s), supporting Hypothesis 3 (see Figure 13). Additionally, the main effect of security framing was significant, $F(1,3196) = 7.80$, $p = .005$, $\eta_p^2 = .002$, with those in the safety framing condition making faster decisions ($M = 6.31$ s, $SD = 4.48$ s) than those in the risk framing condition ($M = 6.99$ s, $SD = 6.88$ s). The interaction between time pressure and framing was not significant, $F(1,3196) = 1.38$, $p = .240$, $\eta_p^2 = 000$.

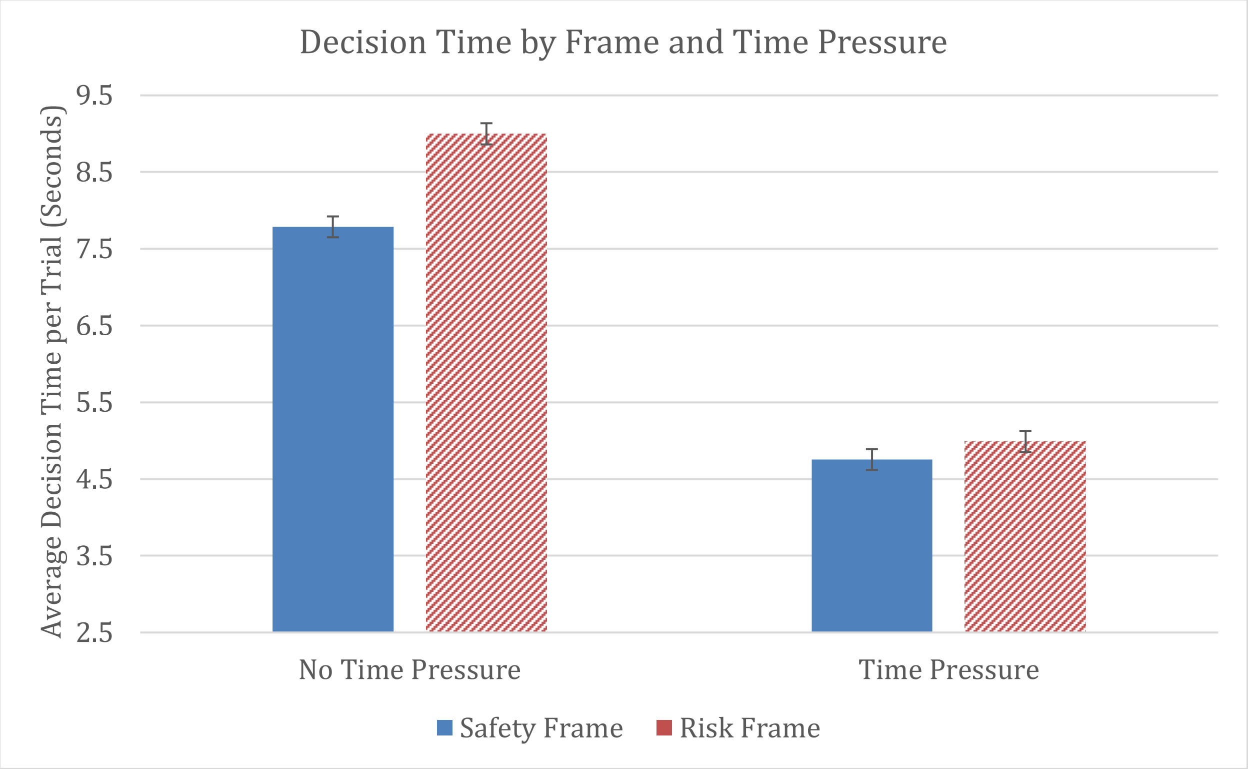


Fig. 13. ANOVA results for decision time (in seconds). The main effects for both time pressure and security framing significantly impacted decision time, while the interaction between the two factors was not significant. Error bars represent standard error.

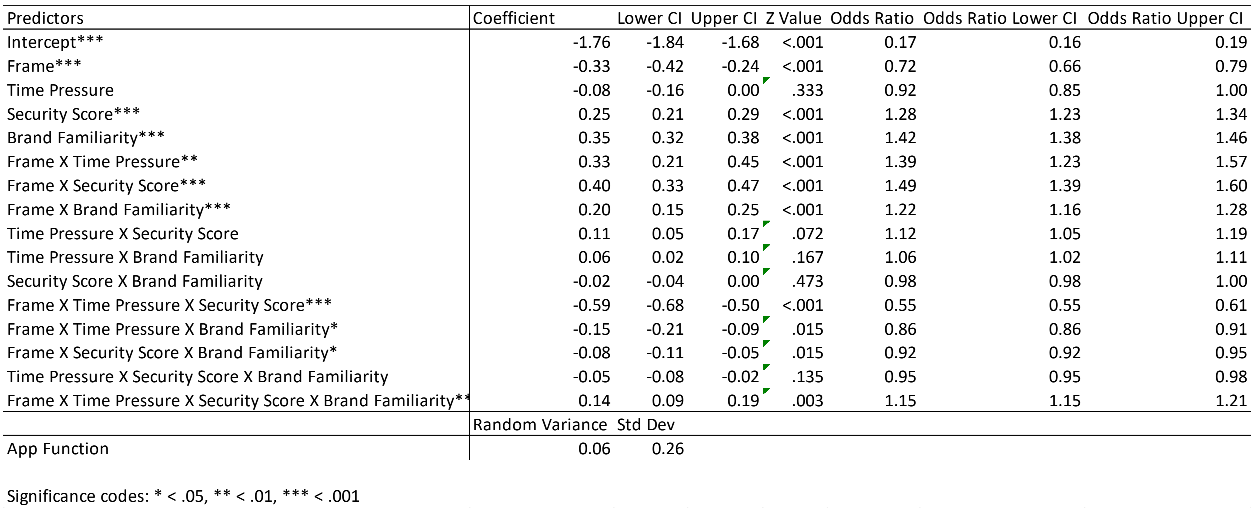
3.2.3 Rationale and Perception Data. Participants' rationale for app choice reported in the post-experiment survey was analyzed using Cochran's Q test. This analysis is similar to a Chi-square analysis but allowed participants to choose multiple responses (brand familiarity, security ratings, user ratings, icon look and feel, other). The Cochran's Q test revealed a significant difference in rationale for app choices, $\chi^2(3) = 46.61$, $p < .001$. "Brand familiarity" was selected as the most influential factor of app choice with 45% of participants, "security ratings" accounted for 44% of participants' decisions, "user ratings" was chosen by 38% of participants, 8% of participants indicated that "icon look and feel" was important, while no participants selected the "other" option. A pairwise post-hoc Dunn test with Bonferroni corrections (α = .016) was conducted to further investigate the differences between rationale choices; only “icon look and feel” differed significantly from the other choices (brand familiarity, security ratings, user ratings; *p*s < .001), whereas all other pairwise comparisons were not significant (brand familiarity vs. security ratings, *p* = .80; brand familiarity vs. user ratings, *p* = .27; security ratings vs. user ratings, *p* = .39).

An attention check in the form of an open-ended question asked participants what the locks below the apps meant. While a majority of the participants (83) provided an answer that was close to the description (answers that included any of the following terms: security, privacy, risk, safety), the remaining 45 provided incorrect answers or failed to understand the meaning of the question (e.g., “yes”, “locks”, etc.). Ideally, all of the participants would have provided a sufficient answer to ensure they all understood the meaning of the locks; however, a Chi-square analysis revealed that the number of incorrect responses was significantly larger than expected (i.e., zero incorrect responses), χ^2 = 158.128.95, *p* < .001.

A concern about the above analyses was that inclusion of results from so many participants that missed the attention check question in the post-experiment survey. To address this concern, another GLMER (see Table 4 for results) and set of LRTs (see Table 5 for results) were conducted with choice data for the 83 participants that passed the attention check. These analyses indicated changes in statistical significance between the two models for time pressure, the interaction between time pressure and security, and the four-way interaction between time pressure, frame, security score, and brand familiarity, none of which were hypothesis-based. Where time pressure was a significant predictor of choice with all participants (*p* = .018), it was not significant when examining only the 83 participants in the follow-up analysis (*p* = .059), meaning that there was no difference in app choice when only considering the absence or pressure of time pressure. Alternatively, where the original analysis indicated that the interaction between time pressure and security scores was not significant (*p* = .052), the follow-up analysis revealed a significant interaction (*p* = .004). This means that, collapsing across risk and safety frames, security scores played a significantly larger role in choice of apps under time pressure than without time pressure. Finally, while the four-way interaction between time pressure, frame, security score, and brand familiarity was significant with all participants (*p* = .006), the interaction was not significant when only examining the 83 participants that passed the attention check question (*p* = .054).

Table 4.

Coefficients in the Adjusted General Linear Mixed Effects Regression

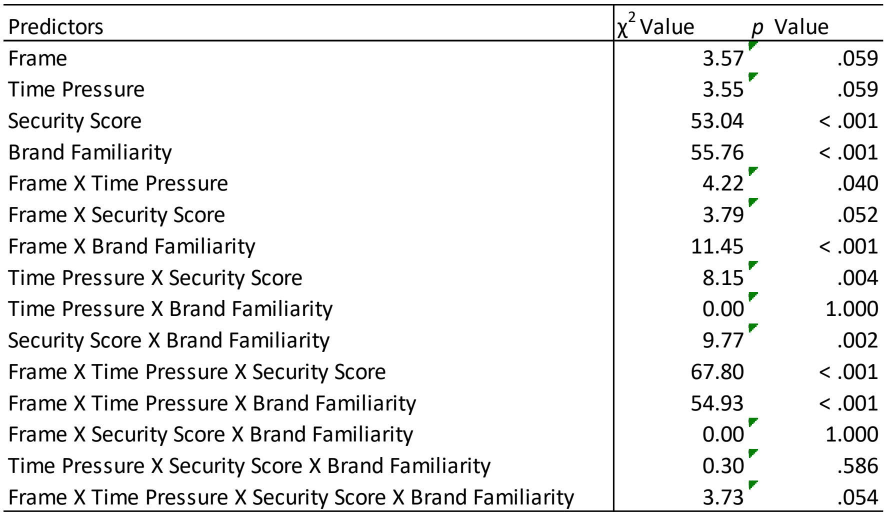


Predictors	Coefficient	Lower CI	Upper CI	Z Value	Odds Ratio	Odds Ratio Lower CI	Odds Ratio Upper CI
Intercept***	-1.76	-1.84	-1.68	<.001	0.17	0.16	0.19
Frame***	-0.33	-0.42	-0.24	<.001	0.72	0.66	0.79
Time Pressure	-0.08	-0.16	0.00	.333	0.92	0.85	1.00
Security Score***	0.25	0.21	0.29	<.001	1.28	1.23	1.34
Brand Familiarity***	0.35	0.32	0.38	<.001	1.42	1.38	1.46
Frame X Time Pressure**	0.33	0.21	0.45	<.001	1.39	1.23	1.57
Frame X Security Score***	0.40	0.33	0.47	<.001	1.49	1.39	1.60
Frame X Brand Familiarity***	0.20	0.15	0.25	<.001	1.22	1.16	1.28
Time Pressure X Security Score	0.11	0.05	0.17	.072	1.12	1.05	1.19
Time Pressure X Brand Familiarity	0.06	0.02	0.10	.167	1.06	1.02	1.11
Security Score X Brand Familiarity	-0.02	-0.04	0.00	.473	0.98	0.98	1.00
Frame X Time Pressure X Security Score***	-0.59	-0.68	-0.50	<.001	0.55	0.55	0.61
Frame X Time Pressure X Brand Familiarity*	-0.15	-0.21	-0.09	.015	0.86	0.86	0.91
Frame X Security Score X Brand Familiarity*	-0.08	-0.11	-0.05	.015	0.92	0.92	0.95
Time Pressure X Security Score X Brand Familiarity	-0.05	-0.08	-0.02	.135	0.95	0.95	0.98
Frame X Time Pressure X Security Score X Brand Familiarity**	0.14	0.09	0.19	.003	1.15	1.15	1.21
	Random Variance	Std Dev					
App Function	0.06	0.26					

Significance codes: * < .05, ** < .01, *** < .001

Table 5.

Coefficients in the Adjusted Likelihood Ratio Tests



Predictors	χ^2 Value	*p* Value
Frame	3.57	.059
Time Pressure	3.55	.059
Security Score	53.04	< .001
Brand Familiarity	55.76	< .001
Frame X Time Pressure	4.22	.040
Frame X Security Score	3.79	.052
Frame X Brand Familiarity	11.45	< .001
Time Pressure X Security Score	8.15	.004
Time Pressure X Brand Familiarity	0.00	1.000
Security Score X Brand Familiarity	9.77	.002
Frame X Time Pressure X Security Score	67.80	< .001
Frame X Time Pressure X Brand Familiarity	54.93	< .001
Frame X Security Score X Brand Familiarity	0.00	1.000
Time Pressure X Security Score X Brand Familiarity	0.30	.586
Frame X Time Pressure X Security Score X Brand Familiarity	3.73	.054

Privacy concerns were also measured via an open-ended question. Over half of the participants (66) expressed specific concerns (e.g., misuse of personal data, unnecessary permissions, tracking location, accessing camera), 6 participants expressed general concerns of privacy (responses such as “yes”), 28 participants expressed no concerns, and 28 participants either did not respond or did not respond in a meaningful manner. Finally, perceptions of the security locks were also measured via an open-ended question. The feedback was generally positive, with several responses complimenting the design overall. Notably, a few of those assigned to the risk frame expressed potential confusion with the design, such that more locks could be misconstrued as greater safety. One participant’s feedback in particular hinted at the stimulus-stimulus compatibility principle (De Houwer, 2003; Kornblum, Hasbroucq, & Osman, 1990), “I think it’s counter-intuitive and confusing. You should instead [use] closed locks and have more locks [equals] better. People are used to more meaning better. Not worse. You have more stars [equals] better and right underneath, more locks [equals] worse. It’s just plain a terrible idea. People are going to think that more locks [equals] better.” Conversely, those in the safety frame described the locks as “intuitive” and a few participants expressed that the design helped them make decisions quickly. Some participants discussed how they would use the security system, “I do like the security lock designs. Even when dealing with brands I knew, it helped remind me of the flaws [inherent] in the brands security. Facebook was a prime example of that.” Other participants seemed less interested in the scores, “App locks do not [affect] my app choices or concerns. If I like it, I keep it. If I don’t like the app, I delete it.”

CHAPTER 4

GENERAL DISCUSSION

The current study focused on the effect of security framing, time pressure, and brand familiarity on user download behaviors of mobile apps. These factors have been shown to individually affect purchase behaviors of other products but had not yet been investigated for mobile apps. As such, the current study was the first to investigate the combination of the framing effect of security scores for mobile apps (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018), time pressure (Madan et al., 2015; Saqib & Chan, 2015; Young et al., 2012), and the effect of brand familiarity (Baker et al., 1986; Harris et al., 2016).

A necessary first step was accomplished through the pilot study, quantifying brand familiarity for mobile apps. While the pilot study provided insight into the perceived familiarity of 300 mobile apps, split into 25 categories, it also examined participants' ratings of favorability and trustworthiness. By adding the latter two constructs, the current pilot study went beyond the creation of a new measure and examined the correlations between the three constructs. The strong correlations between each of the constructs indicates a possible overlap in the belief that a familiar app is both likeable and trustworthy.

Past research on mobile app security scores has shown that such a system can benefit mobile device users and that certain design considerations are more beneficial than others (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018). The main experiment in the current study advanced this line of research by introducing color-coded locks (Rajivan & Camp, 2016) and examining external factors such as brand familiarity (Baker et al., 1986; Harris et al., 2016) and time pressure (Saqib & Chan, 2015; Young et al., 2012). Indeed, the current study provides further support for the security system proposed by Chen et al. (2015), with safety framing resulting in significantly faster and nearly significantly more cyber-conscious decisions than the risk framing. In addition to the behavioral measures of decision time, the subjective reports from participants also supported the use of the safety framed locks. Those in the safety framed conditions praised the locks' intuitive design while those in the risk framed conditions commented on the confusing design, stating that it would make more sense for increasing locks to mean a safer app. This could be due to the confusing nature of the unlocked locks, as locks typically represent a mental model of safety (Rajivan & Camp, 2016). While the unlocked locks were designed to be as equitable to the locked locks as possible, they may have confounded the understanding of the risk scores themselves.

By introducing time pressure, the current study advances both the specific research of mobile app security and the more general research of the framing effect under time pressure. In addition to making faster decisions under time pressure (Madan et al., 2015), mobile users appeared to change the approach by which they make decisions under time pressure, supporting past research that found a reversal in risk preference under time pressure (Saqib & Chan, 2015). Indeed, while participants under the safety frame made safer decisions without time pressure, those under time pressure made much riskier decisions, choosing apps with lower safety ratings. In contrast, those under the risk frame made safer decisions under time pressure than participants under the risk frame that were not under time pressure. These results indicate that, rather than participants becoming globally more risk seeking (Chandler & Pronin, 2012; Madan et al., 2015) or risk avoidant (Ben Zur & Breznitz, 1981; El Haji et al., 2016), there was a reversal in risk preference, as was found by Saqib and Chan (2015). Because the methodology between the current study and that by Saqib and Chan are vastly different, it is not possible to directly compare the reversal in risk preference; that said, the reversal of risk preference under time pressure in the current study lends credence to the notion that the prospect theory’s (Tversky & Kahneman, 1981) S-curve inverts.

The effect of brand familiarity strongly predicted app choice, supporting past literature on other products (Baker et al., 1986; Harris et al., 2016). This result means that mobile users are likely to download familiar apps, along with any potential dangers associated with such apps. This finding, without context, could cause alarm that brand familiarity undercuts the effectiveness of the security score as it stands. However, the current study also showed that, as brand familiarity increased, so too did the differences in downloads between apps with varying security scores. Remember that this is the opposite finding of what was expected; according to Hypothesis 6, it was expected that familiar apps would be viewed as inherently safe, which would negate the perceived need for a security score. Indeed, high-familiarity apps were more scrutinized along the security scale than low-familiarity apps, meaning that participants were choosing safe apps far more than hazardous apps when they were familiar compared to when they were unfamiliar. Note that this finding is likely the result of a floor effect wherein participants were unlikely to download unfamiliar apps by default, which would leave little variability in the preferred security scores for such apps. While this means that participants still relied on brand familiarity by avoiding the apps they did not know, participants also heeded the security scores. If the correlation between familiarity and perceived trustworthiness found in the pilot study is accurate, it appears that participants were relying on this misplaced trust in an app’s familiarity to narrow the options and then referring to the security system among those apps that met the subjective threshold. Therefore, in real-world usage, we can expect that users will gravitate toward familiar apps; that said, it is encouraging that the proposed security system is somewhat robust against the implicit trust correlation found in the pilot study.

The current study showed that time pressure has no effect on how brand familiarity impacts app choice. While this lack of effect means that users are less likely to change their brand preference under time pressure, this may be due to a high baseline of users choosing familiar brands without time pressure. If this is the case, time constraints are unlikely to push people even more toward the apps with which they are familiar.

Beyond measuring the behavioral responses to the experiment, the current study also gathered data on participants' choice rationale, as well as more general attitudes and perceptions regarding cybersecurity and the security score system in question. A substantial number of participants expressed concern for their data and the permissions that apps may request, indicating a potential increase in consumer awareness of the potential threats associated with mobile apps compared to past research (Benton, Camp, & Garg, 2013; Chin, Felt, Sekar, & Wagner, 2012; Felt et al., 2012; Kelley et al., 2012). Furthermore, participants' suggestions and feedback regarding the security system supported the behavioral measures collected. These findings should be considered for future research and design for the security system in question. Likewise, the support for the aforementioned hypotheses suggests that brand familiarity and time pressure play a role in mobile app download behaviors and should be considered in future research and security score design.

4.1 Theoretical Implications

In line with other research on the framing effect, our findings suggest that people tend to be less risk taking when information is framed positively (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981). Indeed, participants that interacted with the safety frame, with the focus on potential gains rather than potential losses, made safer decisions in faster times (Madan et al., 2015). Furthermore, the findings from the study indicate that, under time pressure, the framing effect reverses, as proposed by Saqib and Chan (2015). These findings are critical to better understanding the effect of time pressure on the framing effect.

Beyond examining the framing effect, the current study also showed that brand familiarity has a strong impact on purchase intentions (Baker et al., 1986; Harris et al., 2016). Conversely, there was no evidence to suggest that brand familiarity and time pressure interact, meaning that mobile users are likely to prefer familiar apps regardless of external pressures. This supposed ceiling effect of brand familiarity suggests support for the correlative analysis in the pilot study between brand familiarity, favorability, and perceived trustworthiness, a first in the literature on brand familiarity. This finding is novel and should be further investigated. Based on the expectation that these constructs are intrinsically intertwined, it was expected that brand familiarity, if high enough, would negate the effectiveness of the security scores. This was not the case and suggests that, while users prioritize downloading familiar apps, they still want to make relatively safe decisions. Because users chose familiar apps so much more frequently than unfamiliar apps, it can be expected that the increased sensitivity to the security scores with increasing familiarity is mainly due to a broader pool of security scores from which they would choose. Put another way, if a user were to download an unfamiliar app, the only indication they are making a good choice is through the security score; alternatively, users are likely to narrow their search down to a number of apps with which they are comfortable downloading and use the security scores to further pare down the options. If this is the process by which users truly download apps or make more generic purchases, the implications can guide future research.

Finally, by creating the first brand familiarity Likert scale, we can expect further developments to the current vein of research on brand impact.

4.2 Practical Implications

The current study builds on the prior literature on mobile app security communication (Chen et al., 2015; Rajivan & Camp, 2016; Chong et al., 2018) and supports the use of framing security positively for ease of use. Furthermore, the use of locks resulted in positive comments and subjective evaluations by participants and is recommended for future designs. For systems that implement the safety framing, the closed locks represented the mental models of security well in the minds of the participants, according to previous research (Rajivan & Camp, 2016). Conversely, the current use of unlocked locks in the risk frame may have caused confusion because it used similar iconography. By confounding the concept of risk with a symbol commonly associated with security, the use of locks for a risk frame is not recommended. Instead, iconography that better represents danger can be used, such as crossbones, or X’s. Regardless, the use of a safety-framed system in a real app store has the potential to curb unnecessary risk by users by guiding them to choose safer apps, especially when the apps are relatively unfamiliar. Another finding from the study indicates that time pressure, via external factors or sales/promotions by developers, could result in participants relying less on the security system if framed as safety. However, the positive effect of the safety frame without time pressure is strong enough for us to recommend that only safety framing be used, since participants were quick to point out the confusing nature of a risk-framed security system.

We can expect that in real app stores, familiar apps are likely to be downloaded far more than unfamiliar apps. While an obvious point, it is important to note because participants in the study were more discriminatory among familiar apps along the security score spectrum.

Therefore, the use of the security score in a real app store environment could help users choose an app once they have narrowed their search down to a handful of alternatives. If the brand familiarity scores were then used to compile the most familiar apps for direct comparison along security, this may further serve users.

4.3 Limitations and Future Directions

While the current study is the first to examine the effects of brand familiarity and time pressure on mobile app choices with security scores, there are a few limitations. The pilot study's correlations between familiarity, favorability, and trustworthiness should be further examined in another study. The current results might be due in part to anchoring since the constructs were not randomly ordered nor reverse-coded. Anchoring is the process by which decision makers are influenced by an original choice, becoming less likely to deviate from that "anchor point" (Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1974). For example, if someone were to choose a familiarity score of four, they are then anchored to that score for both the favorability and trustworthiness scales and unlikely to choose scores such as one or seven since they are far away from the anchor point. Therefore, an answer for trustworthiness may have been influenced by the original score for familiarity and favorability.

The attention check in the main experiment failed to garner enough attention to be effective, having only been chosen by 18 of the participants (less likelihood of being chosen than by chance alone). Because of this, we cannot guarantee that all the included data are the result of attentive participation. Likewise, the catch trial in the post-experiment survey was an open-ended question and the resulting responses indicated a lack of understanding of the question asked. Both the attention check and catch trial in the main experiment may have been adversely affected by the design of the risk conditions, with unlocked locks as an indicator of danger. Future research should examine other symbols for the risk conditions to better bolster the impact of the negative frame so that the comparison against the lock image for the safety frame can be more fair.

Another limitation of the study was rooted in the platform used; for those in the time pressure condition, the size of the stimuli may have extended beyond the bottom of the screen, depending on the physical size of the monitor. Participants were able to scroll in order to see the entire trial, but there was no way to equate the relative survey size across all screen sizes. Finally, due to experimenter error, the post-experiment survey did not include a question regarding perceived time pressure. This limitation is less of a concern due to the strong behavioral evidence provided in the above analyses but is nevertheless noteworthy.

Future research should further investigate the correlation between the constructs of familiarity, favorability, and trustworthiness of brands, whether mobile apps or more general products. Given the strong effect of brand familiarity and its interaction with the security system, further research is needed on the effectiveness of the security system when users are interacting exclusively with highly familiar apps. Likewise, the effect of time pressure on the security system and app choice should be further examined, perhaps with varying scenarios, as with sales or out of necessity (e.g., downloading an app to accomplish an immediate goal).

REFERENCES

Aggarwal, P., & Vaidyanathan, R. (2003). Use it or lose it: Purchase acceleration effects of time-limited promotions. *Journal of Consumer Behaviour: An International Research Review, 2*, 393-403.

Baker, W., Hutchinson, J., Moore, D., & Nedungadi, P. (1986). Brand Familiarity and Advertising: Effects on the Evoked Set and Brand Preference. *Advances in Consumer Research, 13*, 637-642.

Ben Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica, 47*, 89-104.

Benton, K., Camp, L. J., & Garg, V. (2013). Studying the effectiveness of android application permissions requests. In 2013 *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 291-296.

Bergum, B. O., & Bergum, J. E. (1981). Population stereotypes: An attempt to measure and define. In *Proceedings of the Human Factors Society Annual Meeting*, 25, 662-665.

Chandler, J. J., & Pronin, E. (2012). Fast thought speed induces risk taking. *Psychological Science*, *23*, 370-374.

Chen, J., Gates, C. S., Proctor, R. W., & Li, N. (2014). Framing of summary risk/safety information and app selection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*, 1461-1465.

Chen, J., Gates, C. S., Li, N., & Proctor, R. W. (2015). Influence of risk/safety information framing on android app-installation decisions. *Journal of Cognitive Engineering and Decision Making, 9*, 149-168.

Chen, J., & Proctor, R. W. (2017). Role of accentuation in the selection/rejection task framing effect. *Journal of Experimental Psychology: General, 146*, 543.

Chen, J., Ge, H., Moore, S., Yang, W., Li, N., & Proctor, R. W. (2018). Display of major risk categories for android apps. *Journal of Experimental Psychology: Applied, 24*, 306-330.

Chin, E., Felt, A. P., Sekar, V., & Wagner, D. (2012). Measuring user confidence in smartphone security and privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 1-16.

Choe, E. K., Jung, J., Lee, B., & Fisher, K. (2013). Nudging people away from privacy-invasive mobile apps through visual framing. In *IFIP Conference on Human-Computer Interaction* (pp. 74-91). Springer, Berlin, Heidelberg.

Chong, I., Ge, H., Li, N., & Proctor, R. W. (2018). Influence of privacy priming and security framing on mobile app selection. *Computers & Security, 78*, 143-154.

Chowdhury, T. G., Ratneshwar, S., & Mohanty, P. (2009). The time-harried shopper: Exploring the differences between maximizers and satisficers. *Marketing Letters, 20*, 155-167.

Clement, J. (2019). Number of available applications in the Google Play Store 2009-2019. Retrieved August 16, 2019, from https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

Clement, J. (2020, January 30). Facebook users worldwide 2019. Retrieved February 18, 2020, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika, 37,* 256-266.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition, 31,* 353-359.

Devlin, J., Ennew, C., McKechnie, S., & Smith, A. (2007). A study of time limited price promotions. *Journal of Product & Brand Management*, *16*, 280-285.

do Prado, R. A. D. P., & Lopes, E. L. (2016). Tick tock, tick tock! An experimental study on the time pressure effect on omission neglect. *Journal of International Consumer Marketing*, *28*, 332-346.

El Haji, A., Krawczyk, M., Sylwestrzak, M., & Zawojska, E. (2016). Time pressure and risk taking in auctions: A field experiment. *Journal of Behavioral and Experimental Economics, 78,* 68-79.

Fagley. N. S., & Kruger, L. (1986). Framing effects on the program choices of school psychologists. Paper presented at the *Annual Meeting of the American Psychological Association*, Washington, DC.

Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012, July). Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 1-14.

Ha, H. Y., & Perks, H. (2005). Effects of consumer perceptions of brand experience on the web: Brand familiarity, satisfaction and brand trust. *Journal of Consumer Behaviour: An International Research Review, 4*, 438-452.

Harris, M. A., Brookshire, R., & Chin, A. G. (2016). Identifying factors influencing consumers' intent to install mobile applications. *International Journal of Information Management*, *36*, 441-450.

Holst, A. (2019). Number of smartphone users worldwide 2014-2020. Retrieved August 16, 2019 from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161-1166.

Jaeger, E. (2014). Facebook messenger: Eroding user privacy in order to collect, analyze, and sell Your Personal information. *John Marshall Journal of Information Technology and Privacy, 31*, 393-421.

Kahneman, D. & Tversky A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47,* 263-291.

Kardes, F. R., Posavac, S. S., Silvera, D., Cronley, M. L., Sanbonmatsu, D. M., Schertzer, S., ... & Chandrashekaran, M. (2006). Debiasing omission neglect. *Journal of Business Research, 59*, 786-792.

Kelley, P. G., Consolvo, S., Cranor, L. F., Jung, J., Sadeh, N., & Wetherall, D. (2012). A conundrum of permissions: installing applications on an android smartphone. In *International Conference on Financial Cryptography and Data Security,* 68-79. Springer, Berlin, Heidelberg.

Kim, H. W., Kankanhalli, A., & Lee, H. L. (2016). Investigating decision factors in mobile application purchase: A mixed-methods approach. *Information & Management, 53*, 727-739.

Klapproth, F. (2008). Time and decision making in humans. *Cognitive, Affective, & Behavioral Neuroscience, 8,* 509-524.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review*, *97*, 253-270.

Laroche, M., Kim, C., & Zhou, L. (1996). Brand familiarity and confidence as determinants of purchase intention: An empirical test in a multiple brand context. *Journal of Business Research, 37,* 115-120.

Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological), 34*, 1-18.

Liu, C. W., Hsieh, A. Y., Lo, S. K., & Hwang, Y. (2017). What consumers see when time is running out: Consumers' browsing behaviors on online shopping websites when under time pressure. *Computers in Human Behavior*, *70*, 391-397.

Madan, C. R., Spetch, M. L., & Ludvig, E. A. (2015). Rapid makes risky: Time pressure increases risk seeking in decisions from experience. *Journal of Cognitive Psychology*, *27,* 921-928.

Maule, A. J., Hockey, G. R. J., & Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, *104*, 283-301.

Mazur, D. J., & Hickam, D. H. (1990). Treatment preferences of patients and physicians: influences of summary data when framing effects are controlled. *Medical Decision Making*, *10*, 2-5.

McGill, T., & Thompson, N. (2017). Old risks, new challenges: exploring differences in security between home computer and mobile device use. *Behaviour & Information Technology, 36*, 1111-1124.

Moore, S. R., Ge, H., Li, N., & Proctor, R. W. (2019). Cybersecurity for android applications: Permissions in android 5 and 6. *International Journal of Human–Computer Interaction*, *35*, 630-640.

Nakashima, R. (2018, August 13). AP Exclusive: Google tracks your movements, like it or not. Retrieved October 11, 2019, from https://apnews.com/828aefab64d4411bac257a07c1af0ecb/AP-Exclusive:-Google-tracks-your-movements,-like-it-or-not.

Ning, R., Wang, C., Xin, C., Li, J., & Wu, H. (2018). Deepmag: Sniffing mobile apps in magnetic field through deep convolutional neural networks. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom),* 1-10.

Park, J., & Stoel, L. (2005). Effect of brand familiarity, experience and information on online apparel purchase. *International Journal of Retail & Distribution Management, 33,* 148-160.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46,* 1023-1031.

Peng, H., Gates, C., Sarma, B., Li, N., Qi, A., Potharaju, R., Nita-Rotaru, C., & Molloy, I. (2012). Using probabilistic generative models for ranking risks of Android apps. In *Proceedings of the 2012 ACM conference on computer and communications security* (pp. 241-252). New York: ACM.

Price, E. (2018, November 26). More than 500,000 people downloaded these malware-infected android apps. Retrieved from http://fortune.com/2018/11/26/google-play-malware-apps/

Rajivan, P., & Camp, J. (2016). Influence of privacy attitude and privacy cue framing on android app choices. In *Twelfth Symposium on Usable Privacy and Security*.

Reyna, V. F., Chick, C. F., Corbin, J. C., & Hsia, A. N. (2014). Developmental reversals in risky decision making: Intelligence agents show larger decision biases than college students. *Psychological Science, 25*, 76-84.

Saqib, N. U., & Chan, E. Y. (2015). Time pressure reverses risk preferences. *Organizational Behavior and Human Decision Processes, 130,* 58-68.

Sawers, P. (2018, November 5). Symantec acquires Appthority and Javelin Networks to bolster its mobile and enterprise security products. Retrieved from https://venturebeat.com/2018/11/05/symantec-acquires-appthority-and-javelin-networks-to-bolster-its-mobile-and-enterprise-security-products/.

Schuster, D., Still, M. L., Still, J. D., Lim, J. J., Feria, C. S., & Rohrer, C. P. (2015). Opinions or Algorithms: An Investigation of Trust in People Versus Automation in App Store Security. In *International Conference on Human Aspects of Information Security, Privacy, and Trust* (pp. 415-425). Springer, Cham.

Shehryar, O. (2008). The effect of buyer's gender, risk-proneness, and time remaining in an internet auction on the decision to bid or buy-it-now. *Journal of Product & Brand Management, 17,* 356-365.

Soliman, M. A. (2017). *The impact of scarcity message on impulsive purchase intention among smartphone shoppers* (Doctoral dissertation, Trident University International).

Stefanko, L. (2019, October 24). Tracking down the developer of Android adware affecting millions of users. Retrieved February 18, 2020, from https://www.welivesecurity.com/2019/10/24/tracking-down-developer-android-adware/

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Vidas, T., Christin, N., & Cranor, L. (2011). Curbing android permission creep. In *Proceedings of the Web, 2*, 91-96.

Vlašić, G., Janković, M., & Kramo-Čaluk, A. (2011). Information hunt: the impact of product type and time pressure on choice of information source for purchase decisions. *Management: Journal of Contemporary Management Issues*, *16*, 87-103.

Wegier, P., & Spaniol, J. (2015). The effect of time pressure on risky financial decisions from description and decisions from experience. *PloS one*, *10*, e0123740.

West, S. G., Ryu, E., Kwok, O. M., & Cham, H. (2011). Multilevel modeling: Current and future applications in personality research. *Journal of personality*, *79*, 2-50.

Wong, Q. (2019). Facebook will reportedly be fined a record $5 billion over privacy mishaps. Retrieved August 16, 2019 from https://www.cnet.com/news/facebook-will-reportedly-be-fined-a-record-5-billion-over-privacy-mishaps/

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology, 59*, 555-561.

Xu, R., Frey, R. M., Vuckovac, D., & Ilic, A. (2015). Towards understanding the impact of personality traits on mobile app adoption-a scalable approach. Paper presented at the *Twenty-Third European Conference on Information Systems (ECIS),* Münster, Germany.

Young, D. L., Goodie, A. S., Hall, D. B., & Wu, E. (2012). Decision making under time pressure, modeled in a prospect theory framework. *Organizational Behavior and Human Decision Processes, 118,* 179-188.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*, 133-148.

APPENDICES

APPENDIX A. LIST OF APPLICATION FUNCTIONS

	App Function		
Brand Familiarity (Low-High)	Apartments/Housing	Banking	Browser
1	PadMapper (1.54)	Varo (1.58)	Cheetah Mobile "CM Browser" (1.57)
2	Apartment Guide (1.85)	GoBank (1.64)	Aloha Browser (1.62)
3	Zumper (1.96)	Frost Bank (1.67)	Brave Private Browser (1.62)
4	HotPads (2)	BankMobile (1.72)	CloudMosa, Inc. "Puffin Web Browser" (1.65)
5	Apartment Finder (2.22)	FirstBank (1.79)	Cake (1.65)
6	ForRent.com (2.22)	Current (1.8)	Mobile_V5 "Web Browser" (1.73)
7	Apartment List (2.39)	Dave (1.85)	Geometry OU "Kiwi Browser" (1.82)
8	Rent.com (2.73)	Ally (2.29)	DuckDuckGo (2.14)
9	Realtor.com (2.93)	Chime (2.6)	Opera (2.32)
10	Trulia Rent (3.24)	Chase (4.25)	Microsoft Edge (3.69)
11	Apartments.com (3.54)	Citi (4.41)	Mozilla "Firefox" (5.49)
12	Zillow (4.79)	Bank of America (5.08)	Google Chrome (6.6)

	App Function		
Brand Familiarity (Low-High)	Dating	Drawing	Ereader
1	happn (1.62)	ibis Paint X (1.65)	FaultException "Lithium" (1.58)
2	Ifwe Inc. "Tagged" (1.7)	MediBang Paint (1.65)	De Marque "Aldiko Classic" (1.6)
3	Hily (1.93)	ArtFlow (1.73)	Moon+ (1.63)
4	Badoo (1.98)	Beste "Paint Free" (1.8)	ITENSE "FullReader" (1.65)
5	Clover (2.06)	Colorfit "PaperColor" (1.82)	media365 (1.65)
6	Zoosk (2.29)	Infinite Painter (1.86)	ReadEra (1.7)
7	Hinge (2.54)	4Axis "Drawing Desk" (1.87)	eReader Prestigio (1.83)
8	Plenty of Fish (2.56)	Creative APPS "Colorfit" (1.99)	MobiPups+ "eBoox" (1.99)
9	OkCupid (2.9)	Autodesk "SketchBook" (2.4)	Kobo (2.03)
10	match (3.61)	Draw it (2.5)	Obreey Products "PocketBook" (2.26)
11	Bumble (4.11)	Adobe "Illustrator" (3.07)	Barnes & Noble "NOOK" (3.64)
12	Tinder (5.62)	Adobe "Photoshop Sketch" (3.96)	Amazon Kindle (5.45)

	App Function		
Brand Familiarity (Low-High)	Fitness	Food Delivery	Games
1	JEFIT (1.78)	RandomAppsInc "Food Button" (1.61)	MochiBits "Left vs. Right" (2.06)
2	FitOn (1.81)	Waitr (1.62)	Kooapps Games "Pictoword" (2.96)
3	despDev "Home Workout" (1.87)	Bite Squad (1.63)	Top Free Games "Bike Race Free" (3.28)
4	PumpUp (1.98)	foodora (1.64)	RobTop "Geometry Dash" (3.34)
5	VGFIT "Fitness & Bodybuilding" (1.99)	Caviar (1.69)	MetroTrains "Dumb Ways to Die" (3.5)
6	Leap Fitness "Home Workout" (1.99)	Seamless (1.8)	Ketchapp Stack (3.52)
7	Leap Fitness "30 Day Challenge" (2.28)	EatStreet (1.81)	Joy Journey "Piano Games" (3.96)
8	Total Fitness (2.32)	BeyondMenu (2.08)	Big Duck Games "Flow Free" (4.66)
9	Samsung Health (2.4)	Postmates (2.53)	NAMCO "PAC-MAN" (4.95)
10	Google Fit (3.46)	DoorDash (5.93)	SYBO Games "Subway Surfers" (5.13)
11	Under Armour "Calorie Counter" (3.66)	Grubhub (5.95)	Imangi Studios "Temple Run" (5.87)
12	Fitbit (4.91)	Uber Eats (6.07)	Halfbrick Studios "Fruit Ninja" (6.03)

	App Function		
Brand Familiarity (Low-High)	Language	Local Business Reviews	Navigation
1	50LANGUAGES "Learn 50 Languages" (1.44)	Qayiem (1.48)	Sygic (1.55)
2	Ati "Learn 33 Languages" (1.6)	SoftDeluxe "Restaurant Guru" (1.56)	HERE WeGo (1.6)
3	Busuu (1.6)	Flying Code "AroundMe" (1.57)	MapFactor (1.7)
4	Mango Languages (1.61)	Resy (1.58)	TomTom (2.09)
5	Beelinguapp (1.64)	Third Coast Interactive "Your Reviews" (1.59)	Voice Navigation Apps "GPS, Maps Driving" (2.18)
6	Memrise "Learn Languages" (1.66)	zomato (1.8)	Maps, GPS Navigation "Offline Maps" (2.33)
7	Language Drops (1.7)	Yell (1.94)	VirtualMaze "Offline Map Navigation" (2.38)
8	HelloTalk (1.78)	Zomato "Urbanspoon" (2.06)	Video Downloader "GPS Navigation" (2.56)
9	Babbel (2.4)	Foursquare (2.39)	GPS Maps Navigation "Navigation & Maps" (2.96)
10	Simon & Schuster "Pimsleur" (2.41)	Yellow Pages (2.92)	MapQuest (3.1)
11	Duolingo (4.28)	TripAdvisor (4.84)	Waze (5.08)
12	Rosetta Stone (4.5)	Yelp (5.32)	Google "Maps" (6.31)

	App Function		
Brand Familiarity (Low-High)	Messaging	Money Transfer	Music
1	Color Cube Studios "Color Messages" (1.73)	WorldRemit (1.63)	Free music player creator "Free Music" (1.59)
2	Tomato 5% Studio "AI Message" (1.8)	TransferGo (1.63)	Free Music - Music Play "Free Music" (2.17)
3	Gather Media "Messages" (1.86)	WigWag (1.64)	Mobile_V5 "Music Player" (2.28)
4	DC Mobile Dev Team "Messaging Classic" (2.03)	Mezu (1.65)	Audiomack (2.46)
5	Best Free Video Editor "Go SMS Pro" (2.44)	Sharemoney (1.71)	Samsung Music (2.65)
6	Contacts Plus team "Messages + SMS" (2.51)	Xoom (1.94)	Music Player. "Music Player" (2.65)
7	Verizon "Messages" (2.99)	Prodoge (3.07)	Google Play Music (4.3)
8	Google "Messages" (3.4)	Zelle (4.4)	Google "Youtube Music" (5.01)
9	Google "Hangouts" (4.49)	Google Pay (4.61)	SoundCloud (5.75)
10	WhatsApp (5.29)	Cash App (5.69)	Pandora (6.13)
11	GroupMe (5.95)	PayPal (5.98)	Apple Music (6.39)
12	Facebook "Messenger" (6.27)	Venmo (5.99)	Spotify (6.42)

	App Function		
Brand Familiarity (Low-High)	News	Notes	PDF Scanner
1	News360 (1.5)	Notas Notepad "BlackNote Notepad" (1.86)	CoolMobileSolution "Fast Scanner" (1.67)
2	SmartNews (1.84)	Notas Notepad "ClearNote Notepad" (1.87)	doo GmbH "Scanbot" (1.73)
3	Reuters News (1.93)	Sappalodapps "Notepad" (1.88)	FireeApps "Tiny Scanner" (1.8)
4	TopBuzz (1.93)	Jacob Ras "Notes" (1.91)	KunKunSoftware "Camera To PDF" (1.89)
5	Particle Media "News Break" (2.06)	Notes "ColorNote Notepad" (2.19)	Appxy "Tiny Scanner" (1.96)
6	Flipboard (2.32)	Evernote (2.24)	Easy inc. "Simple Scan" (1.97)
7	AP News (2.43)	atomczak "Notepad Free" (2.37)	The Grizzly Labs "Genius Scan" (2.07)
8	NPR News (3.22)	Samsung Notes (2.63)	INTSIG "Camscanner" (2.1)
9	HuffPost (4.07)	Google Keep (2.69)	HappyLife Studios "Smart Scan" (2.26)
10	BBC News (4.56)	Lemon, Inc. "Notepad" (3.51)	Smart media "Scanner App To PDF" (2.5)
11	NBC News (4.92)	Microsoft OneNote (4.61)	Adobe Scan (3.49)
12	CNN (5.43)	Office "Notes" (4.74)	Microsoft Office Lens (4.64)

	App Function		
Brand Familiarity (Low-High)	Photo Editing	Ridesharing	Shopping (Ecommerce)
1	AndOr Communications "LightX Photo Editor" (1.85)	NavMake Apps "Lujo" (1.53)	Hollar (1.73)
2	Linerock Investments "Photo Lab Picture Editor" (1.86)	Via "ViaVan" (1.57)	Dhgate.com (2.28)
3	123RF Limited "Pixlr" (1.88)	RideShark Corporation "gobyRide" (1.6)	Jet.com (2.33)
4	InShot Inc. "Photo Editor, Filters" (1.92)	DiDi Global "DiDi-Rider" (1.61)	Mercari (2.84)
5	Lyrebird Studio "Photo Editor" (1.93)	TT RideShare (1.62)	Alibaba "AliExpress" (3.34)
6	dev.macgyver "Photo Editor" (1.97)	Carpooling Ridesharing "Poolmyride" (1.64)	Overstock.com (3.85)
7	Zentertain "Photo Editor Pro" (2.07)	Hitch (1.65)	letgo (3.88)
8	InFrame (2.28)	Via (1.67)	OfferUp (3.96)
9	Adobe Photoshop (2.29)	Zify (1.69)	Google Shopping (3.99)
10	InShot Inc. "Photo Editor Pro" (2.46)	Waze Carpool (3.87)	Wish (5.23)
11	Adobe Lightroom (3.18)	Lyft (5.96)	eBay (5.69)
12	Picsart Photo Editor (3.41)	Uber (6.26)	Amazon Shopping (6.54)

	App Function		
Brand Familiarity (Low-High)	Social Media	Travel	Video Streaming
1	TUBBR (1.55)	Skiplagged (1.65)	Future Today Inc "FilmRise" (1.55)
2	KARMA (1.57)	HolidayPirates "TravelPirates" (1.67)	Screen Media Ventures "Popcornflix" (1.58)
3	Elyments (1.57)	Travelzoo (1.84)	Free Movies TV Shows "Free Movies" (1.65)
4	inLinx Social Network (1.63)	Skyscanner (1.97)	iflix (1.72)
5	WildFyre (1.74)	CheapTickets.com (2.23)	Movies Anywhere (1.84)
6	Vero (1.82)	Hopper (3.04)	XUMO (1.9)
7	LinkedIn (5.59)	KAYAK.com (3.68)	Pluto TV (2.62)
8	Tumblr (5.64)	Booking.com (3.79)	Crackle (3.01)
9	Twitter (6.49)	Travelocity.com (4.25)	Tubi (3.36)
10	Facebook (6.57)	Expedia (4.8)	VUDU (3.61)
11	Instagram (6.6)	trivago (4.81)	Hulu (6.51)
12	Snap Inc "Snapchat" (6.65)	Priceline.com (5.06)	Netflix (6.82)

	App Function
Brand Familiarity (Low-High)	Weather
1	weawow weather app "Weather & Widget" (1.83)
2	Weather Underground (2.11)
3	OneLouder Apps "1Weather" (2.33)
4	ACME AtronOmatic "MyRadar Weather" (2.39)
5	Cleaner & Booster & Secure "Weather Forecast" (2.4)
6	Weather by WeatherBug (2.84)
7	Best App - Top Droid "Weather" (2.98)
8	AccuWeather (3.21)
9	smart-pro android apps "Weather forecast" (3.51)
10	Best Weather App "Weather app" (4.26)
11	Yahoo Weather (4.67)
12	The Weather Channel (5.85)

APPENDIX B. POST-PILOT EXIT SURVEY

1. Please indicate the reason for choosing the applications you did.
    1. Brand familiarity
    2. User ratings
    3. Security ratings
    4. Icon look and feel
    5. Other
2. Please indicate the operating system of your personal cell phone:
    1. Android
    2. iOS (Apple)
    3. Other
3. Please indicate your level of cybersecurity expertise:
    1. (Likert scale 1-7)

APPENDIX C. DEMOGRAPHIC SURVEY

Please indicate below the race with which you identify:

1. Asian
2. Black or African American
3. Hispanic or Latino
4. Native American or American Indian
5. Native Hawaiian or Pacific Islander
6. White
7. Other

Please provide your age: (Open-ended)

Please indicate the gender with which you identify:

1. Female
2. Male
3. Other/Do not wish to answer

Is your vision normal or corrected-to-normal? (Yes/No)

Do you identify as an individual with color-blindness? (Yes/No)

APPENDIX D. PILOT PARTICIPANTS' "OTHER" REASONS FOR DOWNLOADING APPS

"The basic idea behind the app itself."

"If it will be useful for me"

"Use"

"Personal use"

"Necessity"

"What I need the app for"

"Conditions in Terms and Service/type of information they collect"

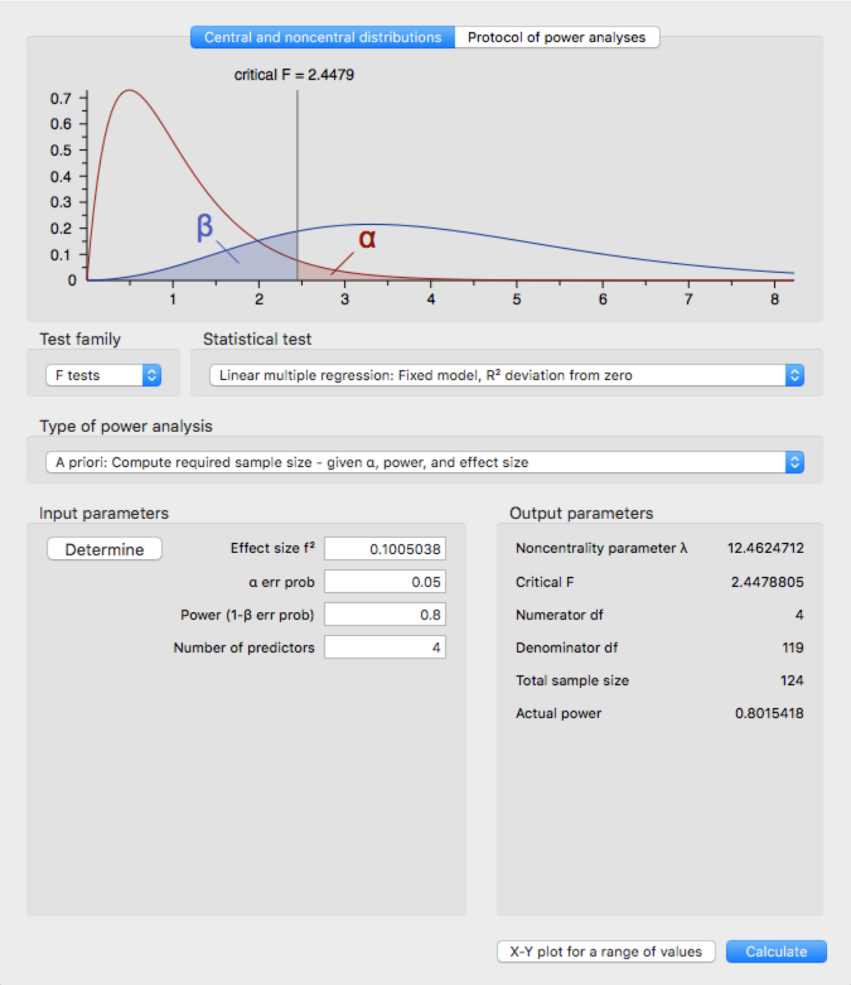
"Recommendations"

"Saw ads for them"

"Advertisements"

APPENDIX E. POWER ANALYSIS FOR MAIN EXPERIMENT

G*Power power analysis using a general linear multiple regression.


Central and noncentral distributions
Protocol of power analyses
critical F = 2.4479
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
β
α
1
2
3
4
5
6
7
8
Determine
X-Y plot for a range of values
Calculate

Test family: F tests

Statistical test: Linear multiple regression: Fixed model, R² deviation from zero

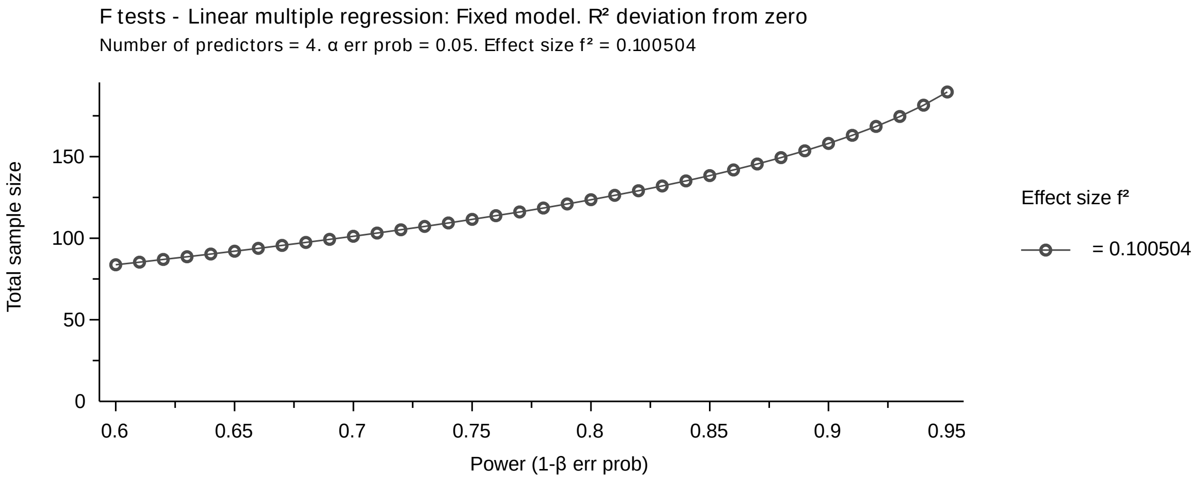
Type of power analysis: A priori: Compute required sample size - given α, power, and effect size

Input parameters

Effect size f²	0.1005038
α err prob	0.05
Power (1-β err prob)	0.8
Number of predictors	4

Output parameters

Noncentrality parameter λ	12.4624712
Critical F	2.4478805
Numerator df	4
Denominator df	119
Total sample size	124
Actual power	0.8015418



APPENDIX F. INTRACLASS CORRELATION

N effective formula to determine minimum participant requirements, given ICC (West et al., 2011).

$N_{\text{effective}} = n_{L1}n_{L2}/(1 + (n_{L1} - 1)\text{ICC})$

level 1 (nL1)	level 2 (nL2)	ICC	numerator	denominator	$N_{\text{effective}}$
144	2	0.021	288	4.003	71.9460405
144	4	0.021	576	4.003	143.892081
144	6	0.021	864	4.003	215.838121
144	8	0.021	1152	4.003	287.784162
144	10	0.021	1440	4.003	359.730202
144	12	0.021	1728	4.003	431.676243
144	14	0.021	2016	4.003	503.622283
144	16	0.021	2304	4.003	575.568324
144	18	0.021	2592	4.003	647.514364
144	20	0.021	2880	4.003	719.460405
144	22	0.021	3168	4.003	791.406445
144	24	0.021	3456	4.003	863.352486
144	26	0.021	3744	4.003	935.298526
144	28	0.021	4032	4.003	1007.24457
144	30	0.021	4320	4.003	1079.19061
144	32	0.021	4608	4.003	1151.13665
144	34	0.021	4896	4.003	1223.08269
144	36	0.021	5184	4.003	1295.02873
144	38	0.021	5472	4.003	1366.97477
144	40	0.021	5760	4.003	1438.92081
144	42	0.021	6048	4.003	1510.86685
144	44	0.021	6336	4.003	1582.81289
144	46	0.021	6624	4.003	1654.75893
144	48	0.021	6912	4.003	1726.70497
144	50	0.021	7200	4.003	1798.65101
144	52	0.021	7488	4.003	1870.59705
144	54	0.021	7776	4.003	1942.54309
144	56	0.021	8064	4.003	2014.48913
144	58	0.021	8352	4.003	2086.43517
144	60	0.021	8640	4.003	2158.38121

APPENDIX G. POST-EXPERIMENT SURVEY

Please indicate the reason for choosing the applications you did.

1. Brand familiarity
2. User ratings
3. Security ratings
4. Icon look and feel
5. Other

What did the security scores mean? (Open-ended)

Do you have any privacy concerns regarding mobile applications? (Open-ended)

Please indicate the operating system of your personal cell phone:

1. Android
2. iOS (Apple)
3. Other

Please indicate your level of cybersecurity expertise: (Likert scale 1-7)

Please provide any additional feedback or suggestions for the experimenters or future designers. (Open-ended)

VITA

Department of Psychology
Old Dominion University
Norfolk, VA 23529

EDUCATION

2018 - 2020 **M.S. (Expected), Human Factors Psychology**, Old Dominion University, VA

Thesis: *The Effects of Security Framing, Time Pressure, and Brand Familiarity on Risky Mobile Application Downloads*

2012 - 2016 **B.A., Psychology**, University of Central Oklahoma, OK

Senior Thesis: *Testing the F-PANAS: Emotion Induction through Film*

SELECT CONFERENCE PROCEEDINGS AND PRESENTATIONS

Mishler, S., **Jeffcoat, C.**, & Chen, J. (2019). Effects of anthropomorphic phishing detection aids, transparency information, and feedback on user trust, performance, and aid retention. To appear in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Mishler, S., **Jeffcoat, C.**, Šabić, E., Yamaguchi, M., & Chen, J. (2019). *Collision or avoidance: Should auditory warnings be toward danger or toward safety?*. Paper presented at the APA: Technology, Mind, and Society Conference, Washington, DC.

Jeffcoat, C., Chen, J., & Proctor, R. (2019). *Influence of verbal protocol requirements on the task framing effects.* Poster presented at the American Psychological Association Conference, Chicago, IL.

Jeffcoat, C. (2017). *Testing the F-PANAS: Emotion induction through film*. Poster presented at the Association of Psychological Science Conference, Boston, MA.

Hamlin, M., **Jeffcoat, C.**, Scott, J., Blaney, P., & Lusey. A. (2016). *The search for an academic intelligence quotient*. Poster presented at the Southwestern Psychological Association Conference, Dallas, TX.