How Task Familiarity and Cognitive Predispositions Impact Behavior in a Security Game of Timing

Jens Grossklags and David Reitter
College of Information Sciences and Technology
The Pennsylvania State University
University Park, Pennsylvania 16802
jensg@ist.psu.edu, reitter@psu.edu

Abstract—**This paper addresses security and safety choices that involve a decision on the timing of an action. Examples of such decisions include when to check log files for intruders and when to monitor financial accounts for fraud or errors. To better understand how performance in timing-related security situations is shaped by individuals' cognitive predispositions, we effectively combine survey measures with economic experiments. Two behavioral experiments are presented in which the timing of online security actions is the critical decision-making factor. The feedback modality in the decision-environment is varied between visual feedback with history (Experiment 1), and temporal feedback without history (Experiment 2). Using psychometric scales, we study the role of individual difference variables, specifically risk propensity and need for cognition. The analysis is based on the data from over 450 participants. We find that risk propensity is not a hindrance in timing tasks. Participants of average risk propensity generally benefit from a reflective disposition (high need for cognition), particularly when visual feedback is given. Overall, participants benefit from need for cognition; however, in the more difficult, temporal-estimation task, this requires familiarity with the task.**

I. Introduction

Security outcomes are often influenced by the human factor. Many decisions are delegated to users because automated security systems have significant limitations, e.g., as a result users are asked to investigate anomalous system behaviors or to assess the trustworthiness of a site or communication partner [6]. However, humans also suffer from predictable weaknesses due to their cognitive limitations and biases [1]; they are limited in their abilities to process information and to correctly implement optimal security strategies. For example, studies have investigated the specific fallacies of decision-making that impact well-being in the future. Humans have a tendency to seek immediate gratification or to procrastinate on important (but costly) decisions [1], [2], [10]. This might contribute to, for example, over-sharing of personal information, or delays with security-relevant system upgrades, respectively.

Less attention has been given to how humans make decisions in continuous time, which shifts the attention from *whether* a particular choice is made to *when* to act. Further, these security choices are typically strategic interactions since attackers learn to optimize their attacks to occur when we least expect them. The optimal timing of security decisions can be important on a long or short time scale. Users have to decide when to conduct backups, when to patch, when to update passwords, or when to check the correctness of personal financial account information. However, humans are also often required to make security decisions with "very little information available, and even less time to be had" [34]. Our work addresses the problem of timing security decisions under the constraints of limited information availability and short decision-making time frames.

Further, identifying the optimal timing of security decisions is typically not without cost. People walk a fine line between performing security actions so often that they take a toll on everyday life, and not performing them often enough or at the wrong times, thus risking, for example, exposure of valuable information or security incidents. The evaluation of when to take an action also requires individuals to take the behavior of attackers into account. To better understand the actions of human decision-makers in such an environment, we conducted two online behavioral experiments based on a specific *security game of timing*, the *FlipIt* game, which we discuss in the following subsection.

A. *FlipIt: A Game of Timing*

FlipIt has been proposed to study the optimal timing of decisions such as cryptographic key rotation, password changing policies, refreshing virtual machines, and cloud auditing [47]. It is specifically inspired by the observation that modern financially-driven cybercrime typically tries to camouflage security compromises for as long as possible to exploit a resource with little interference from the defenders. Attackers may seek such *covert* compromises to increase their gains in high-loss scenarios (e.g., theft of banking credentials to conduct multiple money transfers over time) or low-loss cases (e.g., recruitment into botnets to engage captured resources in multiple spam campaigns). However, covert compromises are particularly harmful if the cost of detecting an attack are non-trivial in comparison to the expected loss from a compromise, and when the assessment whether a resource is secure requires repeated effort. The question is when to take investigative action and to reset the resource to a secure state (if compromised).

The rules of the FlipIt game can be briefly summarized as follows. At each point in time, the game board is *owned* by one of the players. Who is owning the board is not immediately visible to the players, but players can attempt to take over the board at any time. Each such attempt (i.e., a "flip") costs a fixed monetary amount. If a player already owns the board at the time, that price is wasted; if however the opponent owns the board, the player takes over ownership at that time.

A technical description and a motivation of specific design choices are provided in Section III-B.

B. Measuring Individual Differences in Participant Population

In two online behavioral experiments, following the general methodology of experimental and behavioral economics (see, for example, [19]), we investigate the role of individual difference variables in security decision making. In particular, we determine how their general risk-taking propensity interacts with their ability to act successfully in the security scenario [35]. In addition, we study the impact of *need for cognition* which is a measure of "relative proclivity to process information" and "tendency to…enjoy thinking" [8]. Both individual variables were assessed with established psychometric scales [35], [49]. We provide a detailed discussion of the measures in Section III-C.

C. Research Rationale and Objectives

Many security decisions can be improved if individuals pay attention to risk signals, however, such increased readiness comes at a cost. For example, responding carefully to every security warning, indicator or certificate, or reading convoluted privacy policies is infeasible for individuals [6], [16]. Applying dual process theory, we may argue that individuals are limited in their ability to make deliberate and well-reasoned security decisions ("system-2"), and fall back on heuristics that are quick and intuitive, but also error-prone ("system-1") [46]. In our experiments, we cannot directly observe the switch between different cognitive modes of thinking. Instead, we take a somewhat different perspective by asking whether certain innate individual characteristics interact positively or negatively with security decision-making. We approach this research question by complementing the economic experiments with psychometric scales. We then contrast two experiments in which the cognitive cost imposed on an average individual is varied.

Our study is designed to explore the following aspects.

1) Individuals differ in their propensity to make risky decisions. We are interested in how risk propensity affects one's actions in light of changing task familiarity.
2) We suggest that the preference to think deeply (as measured by need for cognition) impacts individuals' performance in the security game of timing. We further expect that the impact is moderated by increased task familiarity.
3) How deliberate (system-2) thinking and decision-making under risk interact remains an open question. As an initial approach to answer this fundamental research question, we explore the interaction effects between risk propensity and need for cognition on the individuals' performance in the security task.
4) Performance likely also depends on the participants' ability to cope with the inherent noisiness of the estimation of the opponent's actions. Further, the cognitive cost of time estimation is likely impacted by the style of visual feedback given to the participants, and whether they have a history of the game available to them. We vary this by playing the game in two different presentation modalities.

In summary, most research on decision-making under risk assumes that riskiness is merely the result of an external random variable, and the subject has a choice between a risky and a less risky strategy. (See, for example, the discussion in [3].) Instead, in a game of timing, we study whether a participant's estimate of when to take action to maximize her payoff is the result of the different variables studied in our experiments. A participant's performance likely depends on her cognitive predispositions, and her ability to predict the opponent's strategy (i.e., the error associated with the timing estimate).

D. Summary and Roadmap

In our study, we investigate how risk propensity and need for cognition impact performance in a security game of timing. We study the impact of these individual characteristics in isolation, but also research how they interact with each other. We further demonstrate how these effects change as decision-makers become more familiar with the task. In total, 456 individuals participated in 6 rounds of two experiments allowing us to base our analysis on over 2500 rounds of decision-making.

Cognitive biases and heuristics can be interpreted as an adaptation to an uncertain environment [28]. They tend to work well in many common situations, but also fail us in many critical, modern-day security contexts. Individual dispositions interact with these more regular biases. Understanding these effects in controlled environments would guide security researchers in studying effects and ways to prevent decision-making failures in practice. Further, we can explore strategies to mitigate these biases through intervention strategies that (in addition to addressing economic incentives [18]) can directly target a cognitive disposition rather than increasing the burden to users through notices and other workarounds [36].

To achieve our objectives, the study effectively combines survey measures with behavioral experimentation. Such academic studies are relatively rare due to the increased complexity of implementation and analysis. However, they allow for more meaningful interpretation of the experimental data and novel trajectories of research (see, for example, an experiment combined with survey measures to understand how self-reported indicators of social capital influence trusting behaviors [15]).

We proceed as follows. In Section II, we discuss related work on games of timing, the FlipIt game, and economic experiments. In Section III, we present our experimental procedure and setup. We present our results in Section IV. We discuss our findings in Section V, and offer concluding remarks in Section VI. The Appendix includes the experimental instructions.

II. Background

A. Security Economics and Games of Timing

Research studies on the economics of security decision-making primarily investigate the optimal or bounded rational choice between different canonical options to secure a resource (i.e., protection, mitigation, risk-transfer), or the determination of the optimal level of investment in one of these security dimensions [21], [25]. These studies have been thoroughly summarized in a recent review effort [30]. In addition, a small number of experimental studies have explored the choice between different security actions, or different levels of security investments [20], [44].

Another critical decision dimension for successfully securing resources is the consideration of *when* to act to successfully thwart attacks. Scholars have studied such time-related aspects of tactical security choices since the cold war era by primarily focusing on zero-sum games called *games of timing* [5]. The theoretical contributions on some subclasses of these games have been surveyed by [41].

Games of timing have found consideration in the economic literature under many different names which include preemption games [42], wars of attrition [23], clock games [7] and stealing games [13].

B. FlipIt: Modeling Targeted Attacks

Our experimental study is motivated by work on the FlipIt game which identifies optimal timing-related security choices under targeted attacks [47]. In FlipIt, two players compete for a resource that generates a payoff to the current owner. Players can make costly moves (i.e., "flips") to take ownership of the resource, however, they have to make moves under incomplete information about the current state of possession of the resource.

The original FlipIt study concerned equilibria and dominant strategies for simple cases of interaction [47]. Other groups of researchers have worked on extensions [40], [31]. For example, Laszka et al. extended the FlipIt game to the case with multiple resources. In addition, the usefulness of the FlipIt game has been investigated for various application scenarios [47]. More recent work investigates the impact on different modeling assumptions about the attacker, i.e., how does the defender's behavior change when faced with different populations of targeting and non-targeting attackers [32], [33].

C. Experiments on Games of Timing

In the domain of experimental and behavioral economics, there has been a renewed interest in non-cooperative games with continuous and asynchronous decision making. The roots of the behavioral research can be found in the 1970s. They concern a variety of games of timing, and duels. For example, Kahan and Rapoport studied duels with symmetric and asymmetric accuracy functions and number of bullets, respectively [26], [27].

Experiments show that continuous time decision-making is challenging for humans and game-theoretical predictions frequently fail to explain experimental observations. For example, Friedman et al. observe that convergence can fail even when iterated deletion of dominated strategies would theoretically lead to the Nash equilibrium [12].

In this paper, we describe and analyze two experiments which draw from the theoretical model of the FlipIt game [47]. Preliminary analysis of the first experiment [37] shows that participant performance improves over time (however, older participants improve less than younger ones). We also found significant performance differences with regards to gender and the need for cognition. Further, we have begun to determine the rational strategies and develop a model backed by a cognitive architecture, which described human heuristics that practically implement risk-taking preference in timing decisions [43]. In the present manuscript, we extend our preliminary work with an in-depth analysis of the experimental data, and the analysis of the data of a second experiment. In particular, we study the interaction effects between the psychometric measures of the desire for deep reasoning about a problem and the general propensity of risk taking with task experience and how those factors explain task performance.

A better understanding of how individuals' abilities and predispositions interact in complex security situations is important to understand past security failures. Studies that merely focus on observed behaviors fail to consider the underlying reasons for why individuals act in certain ways. As such, our work can help to select and eventually train individuals to act more successfully in security scenarios that necessitate quick decision making in continuous time.

D. Online Experimentation

The two experiments were conducted online and used the Amazon Mechanical Turk (AMT) platform, which connects requesters of services (e.g., researchers) with individuals willing to perform tasks. Experiments on AMT are able to reach a large number of potential subjects in a relatively short period of time, for a cost comparatively lower than traditional laboratory studies [24].

While AMT was originally intended to perform tasks that were difficult to automate (e.g. translation; see [9]), the service has since gained popularity and is commonplace in behavioral research including privacy and security studies. For example, Sheng et al. investigated susceptibility to email-based phishing schemes [45], Christin et al. studied individuals willingness to engage in unsafe online behaviors in exchange for payments [10], Wang et al. researched how users engage with privacy configuration interfaces when installing social applications [48], and Nochenson and Grossklags conducted an experiment on consumers' vulnerability to fall for post-transaction marketing scams [38]. Further, despite concerns about the validity of using AMT for research studies, it has been shown that AMT participants "produce reliable results consistent with standard decision-making biases" [17].

In Section III-D, we discuss the specific recruiting practices we followed for our experiments on AMT.

III. Methodology

A. Overview

Our analysis in this paper relies on the data from two experiments which follow the general methodology of experimental and behavioral economics [19]. In contrast to many experiments, we did not bring experimental subjects into a physical laboratory. Instead, we utilized Amazon's Mechanical Turk service to run our online experiments [24]. The experiments were set up much like common laboratory experiments, and were approved by our university's Institutional Review Board.

At the beginning of the study, participants were presented with a consent form which detailed the procedures that they were to follow, the structure of payments, and a number of other pieces of pertinent information. After consenting to the terms of the experiment, participants were redirected to an instructions page. This page stated the rules of the game and also described an example game. Further, we included a detailed description of the payment structure.

Once participants consented to take part in the experiment and read the rules of the game, they were redirected to a survey questionnaire (see Section III-C). After successfully completing the survey, participants were redirected to the main page of the experiment. On this page, we displayed the game "board" which showed the actions and results of the FlipIt game, a button to start a new game round, and a button to "flip" the board. Additionally, participants were given the option to revisit the rules of the game.

B. Implementation of FlipIt

FlipIt is modeled as a finitely-long game with two competing players who each aim to maximize ownership of an indivisible resource (i.e., the *board*). The resource's current state is two-valued (i.e., indicating which player has control of the resource), but covert. Each player has a single constant-cost move (i.e., the *flip*) which has the effect to take/re-take control (if the other player has ownership), or merely to maintain control over the board (if the other player does not have ownership).

Our specific implementation of FlipIt reflects a number of design choices. First, the individual payoff of the players increases linearly with the time they have ownership of the resource. This choice is suitable if we are considering compromise of a networked resource for the purpose of sending spam, but alternative scenarios are plausible as well. Second, we reveal information about the past state of the resource to the player after each flip. That is, at the time of each flip the defender learns whether the resource has been compromised since her previous flip (and she then also learns the exact payoff since the previous flip). However, the current state forward will again be covert. This design choice reflects a middle-ground; we defer to future work the experiment in which defenders only receive feedback about their payoff performance and the state of the resource after the experiment has ended. Further, we would consider the opposite case (i.e., a game without any covert state) less relevant for the security context. Third, our experimental setup focuses on relatively fast-paced games with a length of 20 seconds. With this initial set of experiments, we explore the challenges of timing security decisions that require quick reactions and fast information processing [34]. In future work, we plan to explore also decision-making as it unfolds over medium and long time frames.

C. Survey and Psychometric Scales

The survey consisted of four parts. The first part of the survey asked participants basic demographic information, including their age, gender, level of education, and country of origin. The next three parts of the survey were presented in randomized order. One part was a set of integrity check questions to verify participants' attention to the details of the survey.

The other two sections in the survey were psychometric scales[1] that assessed the level of *risk propensity* (from [35]) and *need for cognition* (from [49]) of participants. Below we briefly present the two psychometric scales.

- The scale for risk propensity (RP) that we utilized consists of 7 questions. RP is a measure of general risk-taking tendencies [35]. Note that the RP measure does not define an absolute measure of risk-neutrality with respect to a rational task analysis. Instead, risk-taking is measured within its sample distribution.
- To measure need for cognition (NFC) we used a 5-question scale. This shortened scale has been tested and verified to be usable as an alternative for the original long scale [49].[2] NFC is a measure of "relative proclivity to process information" and "tendency to…enjoy thinking" [8]. That is, individuals with low NFC are typically not motivated to engage in effortful, thoughtful evaluation and analysis of ideas. As a result, they will be more likely to process information with low elaboration, i.e., heuristically [11]. However, individuals with a high NFC may also still be guided by intuition, emotions, and images, but they will use these factors "in a thoughtful way" [39]. Therefore, the assumption that one can equate a high NFC with fully rational reasoning has to be treated with caution [39]. As a result, the exact impact of high NFC is an empirical question which we investigate in the context of security games of timing.

D. Recruiting and Participant Pool

We restricted the pool of participants to include only Mechanical Turk users based in the United States who had an approval rating of over 90%. We put these restrictions in place to ensure that the subject pool was as minimally influenced as possible by variables other than those of interest (i.e., we did not aim to compare data for different countries of origin) and that there was minimal noise (from participants who frequently had their work rejected for poor quality). The experiment was run as a number of distinct Mechanical Turk Human Intelligence Tasks (HITs). Participants were not allowed to participate multiple times in our experiments, since we were interested in measuring the impact of task experience.

E. Rounds

For the purposes of this experiment, participants played six rounds of the FlipIt game that lasted 20 seconds each. The first round was introduced as a "practice" round. In the practice

[1] Psychometric scales measure psychological constructs, usually with regards to individual differences. The scales that were used in this paper were a series of Likert-style questions that were aggregated to yield a measure of the construct desired (here risk propensity and need for cognition). While there are a number of other ways to measure these constructs (such as the Iowa Gambling Task [4] for Risk Propensity), we felt that introducing other non-survey activities into the process would distract from the task at hand. Therefore, despite the additional confidence that may have been found using these other types of tests, we opted for the simpler and less distracting scales.

[2] The scale is a tested and verified shorter version of the original survey instrument with 34 items [8], [11].

round, participants were not eligible for a bonus payment. The start of the five experimental rounds was signaled to the participants with a pop-up message. The only difference between the practice round and the non-practice rounds is the presence of a possible bonus payment which was awarded to the participants according to their performance. Participants were informed of these rules on the instructions page. Further, we restated these facts to them at the beginning of the practice and experimental rounds, respectively.

Our experimental setup, involved each human participant in a relatively fast-paced version of the FlipIt game. However, we expected that the round length of 20 seconds would provide participants with enough time to develop an appropriate strategy against the computerized opponent. We used five paid rounds to give players enough time to improve their performance in the game and a single practice round in which to experiment without penalty.

F. Participation Fee and Incentive Payments

Participants earned a show-up fee, a, for completing the study irrespective of their performance in the game. They played $n = 5$ experimental rounds numbered $1 ... n$ and a single practice round numbered 0. Participants were paid according to the point difference between the points awarded for their own performance and the points awarded to the computerized opponent for its performance. (Importantly, this means that both net payoffs could be negative and still give a positive outcome for the human or computerized participant.) Let the point difference in round i be known as Δ_i. That is, if a player won by 200 points in round 1 then $\Delta_1 = 200$ and if he loses by 900 points in round 2 then $\Delta_2 = -900$. Let e be the exchange rate for points (i.e., the monetary value of a single point in dollars). Let the per-round endowment be x_i. The purpose of the endowment was to allow participants to experience relative losses.

The payment function for a single participant is as follows:

$$\text{Bonus payment} = e \sum_{i=1}^{n} \max(x_i + \Delta_i, 0) \qquad (1)$$

$$\text{Total payment} = a + \text{Bonus Payment} \qquad (2)$$

The practice rounds are not included in the payments above. We visualize the payment structure in Figure 1, which shows that individual payoff increases linearly with the amount of time the board is held.

In this experiment, we set the exchange rate $e = 0.0001$ which corresponds to 100 points = \$0.01. The per-round endowment was set at $x_i = 1000$ points $\forall\ 1 \leq i \leq n$. I.e., in a tied game the participant would earn \$0.10.

The total payment (Equation (2)) was paid to the participants in two installments. Participants first accepted our HIT that paid \$0.50 upon successful completion. This corresponds to the show-up fee a above. After the experiment was completed, participants were paid a bonus payment through the Mechanical Turk system equal to the remainder of the total payment (i.e., the amount given by Equation (1)).

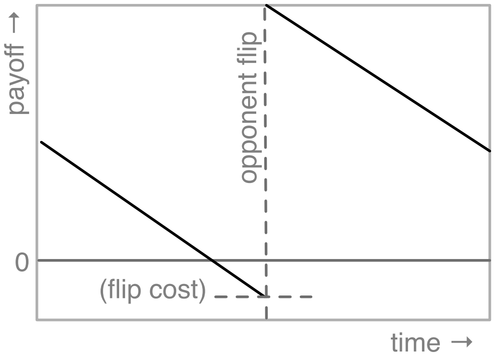


Fig. 1: Player payoff in relation to timing of opponent's flip (at center), for any flip but the first one.

G. Strategy of the Computerized Player

In both experiments, participants faced the same type of computerized opponent that played a fixed (non-adaptive) periodic strategy. The opponent's flip spacing (*tick*) and time of first flip (i.e., *anchor*) changed in every round of the game, but the overall strategy of the opponent did not. (Please note that the human participants always owned the board at the beginning of each round. The computerized player would first act based on the randomly determined anchor value.) Both values were drawn from uniform distributions before a new round started. Flip rates ranged from 1 to 5 seconds, and the anchor ranged form 0.1 to 4.1 seconds.

In the instructions, we did not inform the human participants that they would be paired with a computerized player. Previous research has shown that varying the information about the type of opponent player (i.e., human or computerized) can impact the strategies and outcomes in a competitive game (see, in particular, [22]). We did not explicitly vary such information in our experiment. However, we did vary information about the strategy of the opponent (as described below).

H. Information Treatments

We implemented four separate information treatments which gave successively more information to a participant about the strategy of the computerized opponent. Subjects were randomly assigned to one of the information treatments. For the purpose of the analysis in this paper, we pool the data across the information treatments. (For additional details about the information treatments see [37].)

I. Different Visual Feedback: Experiments 1 & 2

We assigned subjects to one of two different systems of visual feedback, and took measures to prevent double participation.

In Experiment 1, the *visual modality* experiment, subjects saw all past moves (until their most recent flip) on a single timeline (i.e., they had a partial history of the game available). See Figure 2. For the benefit of the reader, we include the instructions page for Experiment 1 in the Appendix.

In Experiment 2, the *temporal modality* experiment, we showed participants only the results of their most recent move.

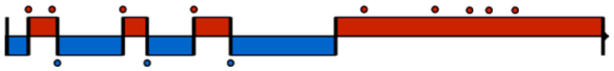


Fig. 2: Feedback shown in the visual modality version (Experiment 1). This information was updated after each flip, so that participants saw data to the left of their last flip. Also shown was a progressing grey bar to the right. Blue dots represent flips by the participant, red dots indicate opponent flips.

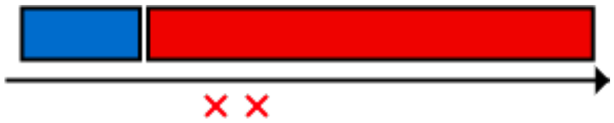


Fig. 3: Feedback shown in the temporal modality version (Experiment 2). This feedback was shown after each "successful" flip that gave the player back control. Blue area is area that was under control by the participant until the red player took control. Each red X indicates that the opponent made a superfluous flip (the position of each X is inconsequential).

That is, they had no visual representation of the history of the game available to them. See Figure 3.

The additional information available in Experiment 1 allowed participants to visually extrapolate distances rather than having to decide on an optimal time to flip without this additional information. In addition, Experiment 2 poses higher demands on individuals' working memory, i.e., they need to try to remember when the opponents' actions happened.

We believe that Experiment 2 requires an initially higher degree of system-1 type decision-making, because individuals have to take intuitive actions to gather data under the more difficult regime with only temporal feedback and without visual access to the game history.

J. Focus of the Analysis

In classical experiments on decision-making under risk, the external random variable is truly random, and the subject has a choice between a risky and a less risky strategy. In a game of timing, the participant's estimate of when to take action to maximize her payoff is the result of different variable factors which we study with our experiments.

1) A participant's performance likely depends on her ability to produce an accurate estimate and her understanding of the task. We address these aspects in two ways. First, we study the impact of individual differences between subjects (as measured by the psychometric scales) on performance in the game. Second, we investigate the role of experience and task learning by allowing the participants to engage in multiple rounds of the game.
2) Performance likely also depends on the inherent noisiness of her estimation. We vary this by playing the game in two different presentation modalities. In Experiment 1, the history of the game is visualized along a horizontal time bar shown on the screen. The time bar continuously extends to the right as the game goes on. This way, the subject is able to visually extrapolate the opponent's flips (which are periodic). In Experiment 2, the history of the game is not visualized, and the subject has to remember the time durations between opponent flips. Thus, subjects use a cognitive, temporal estimation mechanism to pinpoint the opponent's moves. We expected Experiment 2 to be more challenging for individuals.

IV. Results

A. Subjects and Demographics

310 participants completed the first experiment, and 151 participants completed the second experiment. These participants were recruited on the Amazon Mechanical Turk platform and were based in the United States. We excluded 5 subjects from Experiment 1 due to double participation. In total, we utilized the data for 456 participants who played 2736 rounds of the FlipIt game.

320 (70%) of the participants were male, while 136 (30%) were female. The mean age of the participants was 29.5 (sd = 9.9), with less than 15% of the participants being older than 40 years. 45% of the participants had completed "some college" and an additional 45% had obtained at least a four-year college degree.

B. Regression Model

We fitted three regression models to understand the impact of the key variables on the payoff earned by the individuals (see Table I, which shows the minimized, final model after stepwise regression; with significance threshold $p < 0.10$).

In this section, we investigate the aggregate data across both experiments (i.e., the column marked "Combined"). However, the models for data relating to the individual experiments support many of these findings (see columns marked "Exp. 1" and "Exp. 2"). Note that we capitalize variables to increase readability.

As expected, we find that Experiment 2 came with lower monetary prospects for participants, likely due to its increased difficulty. Age had a small negative impact on performance in the experiments (see effect for ln Age), primarily in the later rounds (see interaction effect for (Round):ln Age).

A higher Tick and a larger Anchor increased the payoff of the human participants.[3] The former finding is intuitive since the human participants have to flip less often to maintain control of the board. The latter finding is straightforward since participants maintain longer control of the board at the start of the game.

Increased experience with the game (as measured by Round) improved performance, i.e., participants gained approximately an additional 2 cents in each subsequent round.

We include terms for need for cognition. This variable (abbreviated NFC) is, like others, centered around 0 (mean), with standard deviation of 7.09, and range $[-19.37, 13.63]$.

[3] Tick indicates the computerized opponent's flip spacing, and anchor is the time when the computerized opponent flips for the first time. The human participant had ownership of the board at the start of the game in all rounds. See Section III-G.

Based on exploratory analysis (Figure 6a), risk propensity was first centered around 0 (mean), then transformed by taking the absolute to provide a measure of *deviation from the mean*, and log-transformed. We abbreviate this measure as $\ln|RP|$. Standard deviation of this transformed measure is 1.02, mean 1.72, range $[-0.84, 2.45]$.

In the combined data for both experiments, we find a main effect for NFC indicating improved performance by participants with higher need for cognition characteristics. The combined model finds no main effect for $\ln|RP|$. However, we discuss interaction effects for NFC and $\ln|RP|$ in the following sections.

C. Interaction Effect of Need for Cognition and Task Experience on Performance

Aggregated across both experiments, we find that a higher NFC benefits individuals irrespective of the level of experience (see Figure 4a). This observation is primarily driven by data from Experiment 1 (see Figure 4b). (Note that for *all graphical* representations, we classified the population of subjects as higher-than-average and lower-than-average NFC and RP levels, respectively.) Concretely, a participant with a need for cognition characteristic of one standard deviation above the norm ($sd = 7.1$) achieved 0.8 cents more in payoff per round.

This applies to the situation of average task experience. Task experience matters more in Experiment 2 (see Figure 4c). Here, we find that initially a high NFC is associated with lower payoffs numerically. Only with increased task experience do we observe the superiority of an increased tendency for thinking. While Round is reliably correlated with higher performance in all models ($p < 0.0001$), this performance improvement appears to steepen with higher NFC in Experiment 2 ($p < 0.10$, Table I, and Figure 4c).

Accordingly, NFC has a nuanced impact. We observe that a tendency to think deeply about the game and to arrive at strategy choices through system-2 thinking is often but not always helpful (see Figures 4a - 4c). The interaction effect of NFC and task experience may have been influenced by general task difficulty and the visual feedback that participants received. In the visual modality of Experiment 1 (i.e., the easier game), participants generally benefit from higher NFC levels, both in early and late stages of the task (see Figure 4b). In the temporal modality of Experiment 2, this benefit, if any, is increased with greater task experience; otherwise, intuitive thinking appears to trump analytic reasoning (see Figure 4c).

D. Interaction of Risk Propensity and Task Experience

In both experimental modalities, risk propensity affects performance in a similar way (see Figures 5a-5c). Risk-seeking subjects do better once they have gained experience. In particular, in the final two rounds, subjects appear to benefit from a tendency to seek risks.

As explained in the Section IV-B, a measure of log-transformed deviation of risk propensity is included.

Over the full set of rounds, the full regression model indicates an interaction effect between risk deviation and experience of about 0.13 cents per unit of log-transformed risk deviation. This translates to an additional round profit of 0.23 cents per round played for a participant whose risk propensity is one standard deviation higher or lower than average.

E. Interaction Effect of Risk Propensity and Need for Cognition

Finally, we study the interaction effect between the two individual difference variables on task performance (see Figures 6a-6c). The exploratory analysis suggests that individuals' NFC preferences dictate whether they benefit from risk biases. In the aggregate data and in Experiment 1, we observe that for very risk averse and highly risk-seeking individuals there is no observable correlation of task performance and NFC. In contrast, average risk-seeking individuals benefit strongly from a high NFC, and suffer considerably from a low NFC.[4] This effect is particularly strong for Experiment 1 (see Figure 6b).

The risk deviation $\ln|RP|$ captures the interaction with need for cognition. The models show a reliable effect for the combined data and Experiment 1. Thus, it is not risk propensity per se, but the deviation in risk propensity from the population average that seems to reduce the effect of NFC and increases learning.

V. Discussion

Our experiments require implicit, system-1 type decision-making. In particular, we believe that Experiment 2 requires an initially higher degree of system-1 type decision-making, because individuals have to take intuitive actions to gather data under the more difficult regime with only temporal feedback and without visual access to the game history. Need for cognition is a metric that allows us to assess whether subjects like to engage in system-2 reasoning. Thus, it makes sense that a high NFC is beneficial particularly in Experiment 1 with partial history availability, and less so (or later in the game) in Experiment 2. With this relationship, we show that the survey-based measure of NFC can be predictive of payoff performance. For the economic experimenter, this emphasizes the importance of selecting an adequate subject population.

This becomes even more relevant when considering the role of risk propensity as participants learn to do the task. Risk-seeking personalities fare well once they are ready to analyze the task (Rounds 4 and 5, Fig. 5a, 5b). Before task experience is acquired, this is not the case.

Subjects can, at times, make up for their risk propensity, or moderate their risk-taking in relation to task experience. As we show elsewhere in a psychological experiment [14] using a different, but comparable timing game, subjects with a range of risk propensities can play at a comparable level. There, risk-seeking participants took increased risks primarily in fast games and while inexperienced. In an analysis of the timing in Experiment 2 [43], we show that risk-seekers play early (and risky) at the beginning of the game, but late (and conservative) towards the end. Risk-avoiders do the opposite. We argue that risk-taking preferences can provide a behavioral default whose influence reduces as more experience with the task is acquired.

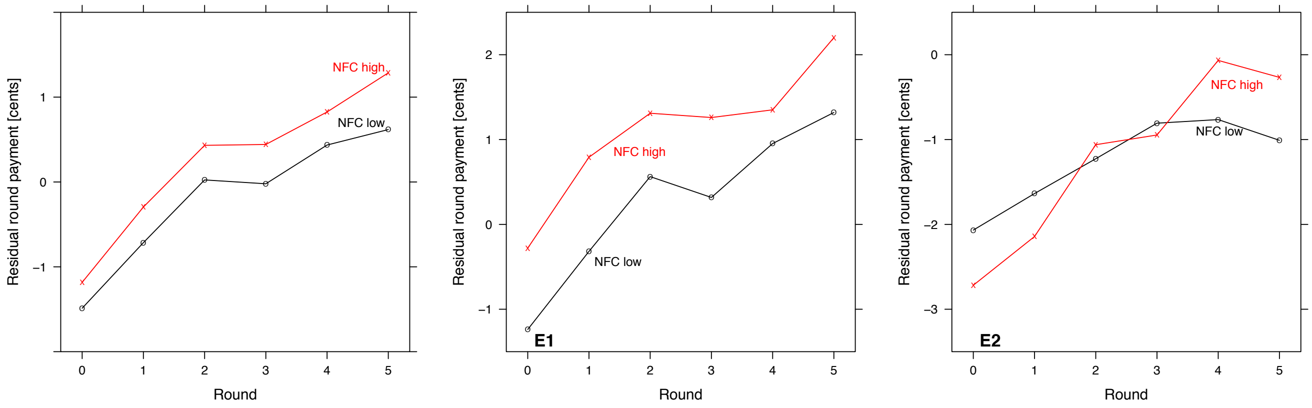
Risk propensity leads to different outcomes for individuals that have different needs for cognition. Figs. 6a and 6b show

[4] Note that we do not define risk-neutrality with respect to a rational task analysis. Instead, risk-taking is measured within its sample distribution.

	Combined	Exp.1	Exp.2
Intercept	10.380^{***}	11.515^{***}	5.312^{***}
Experiment 2	-1.866^{*}		
ln Age	-1.005^{*}	-1.382^{**}	
Tick	0.011^{***}	0.012^{***}	0.010^{***}
Anchor	0.005^{***}	0.005^{***}	0.005^{***}
(Round)	1.986^{***}	1.891^{**}	2.193^{*}
(NFC)	0.110^{**}	0.139^{***}	
(NFC):ln(|RP|)	-0.038^{*}	-0.047^{**}	
(Round):ln Age	-0.536^{***}	-0.473^{*}	-0.622^{*}
(Round):(NFC)			0.021^{+}
(Round):ln(|RP|)	0.130^{**}	0.080^{+}	0.156^{*}
Ω_0^2	0.36	0.37	0.30
Log Likelihood	-8011	-5355	-2666
Deviance	16022	10710	5332
Num. observations	2736	1830	906
Num. subjects	456	305	151

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.10$

TABLE I: Three linear mixed effects regression models. Response variable: round payment in cents. Normalized predictors are shown in parenthesis; they were centered around 0. Tick and Anchor in sec/100, participant's age in years. One random intercept, grouped by subject. Significance levels obtained via t-test, but all levels were confirmed via bootstrapping (95% confidence intervals for ***,**, and *, and 90%-C.I. for $^{+}$, $n_{sim} = 500$). Ω_0^2 indicates variance explained (a substitute of R^2 for mixed-effects models, cf. [50]). All fits obtained using R packages “lme4” 1.1-4, and “LMERConvenienceFunctions” 2.5.



(a) Combined Data

(b) Experiment 1

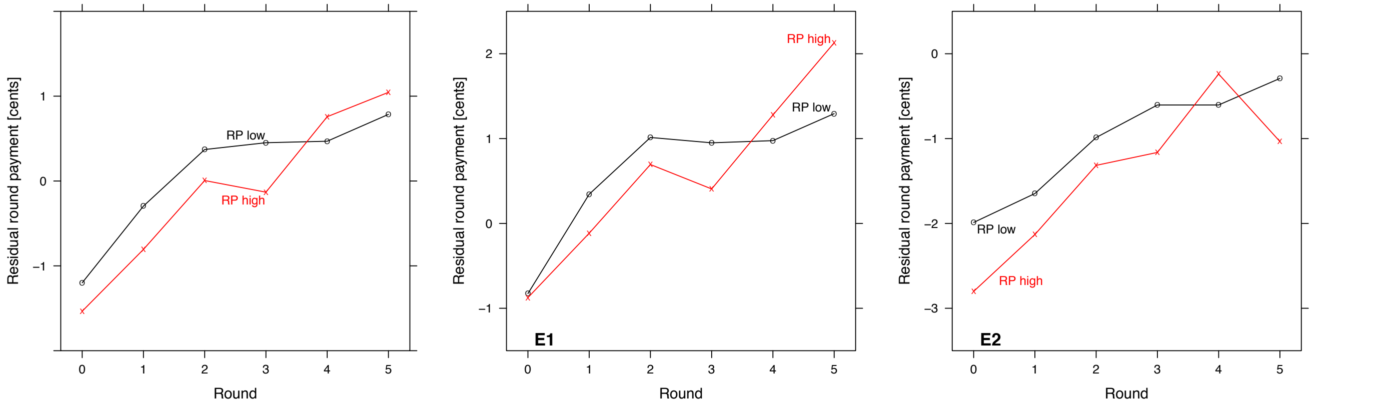
(c) Experiment 2

Fig. 4: Impact of task experience and need for cognition (NFC) on round payoff after regressing out the effects of ln Age, Tick and Anchor. NFC is divided into higher- and lower-than-median values.

that average risk propensity may help those with high need for cognition, and hurt those with low need for cognition. The strength of this effect may be moderated somewhat by the nature of the task and the availability of information used in explicit reasoning. In contrast, in the aggregate data and in Experiment 1, we observe that for very risk averse and highly risk-seeking individuals there is no appreciable correlation of task performance and NFC. However, it is important to keep in mind that these results are correlational.

Performance in the game is impacted by each participant's estimate about the timing of the actions of the opponent (i.e., primarily the observations about the computerized player's flip rate). According to the design of the task, there are two variables that should reduce the noise in this estimation task that induces risk. Familiarity with the task, but also precision of the estimates of opponent's actions. Both of these commonly occur in security scenarios. However, to understand the cognitive process involved and to draw conclusions about policy, it is necessary to differentiate between these two influences. Participants consistently earn higher payoffs with task experience. Experiments with longer rounds than 20 seconds (with about 3-8 flips, as used in the present study) will be needed to better support a learning effect. High need for cognition, as well as particularly high or low risk propensity intensify learning.

As the plots (Figures 5a–6c) show, the interaction between the two individual-difference properties we examine is non-linear. The underlying cognitive processes will be

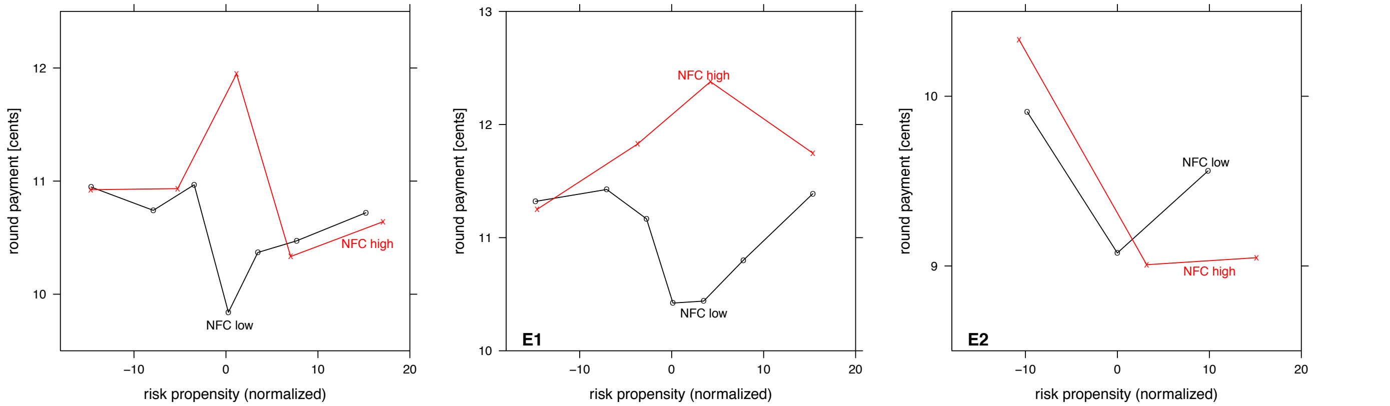


(a) Combined data

(b) Experiment 1

(c) Experiment 2

Fig. 5: Impact of task experience and risk propensity (RP) after regressing out the effects of ln Age, Tick and Anchor. RP is divided into higher- and lower-than-median values.



(a) Combined Data

(b) Experiment 1

(c) Experiment 2

Fig. 6: Impact of need for cognition and risk propensity. NFC is divided into higher- and lower-than-median values.

better reflected in more complex process models than the presented mixed-effects models designed to determine the relevant variables. For instance, the observed data would be consistent with a cognitive model that shifts the optimal risk-taking for system-2 thinking in accordance with the task (see [14] for models that contrast risk-taking and patience in a similar task). Risk-seeking facilitates exploration of new strategies, while risk-avoidance leads to exploitation of the best strategy learned so far. Which one is most appropriate depends on the availability of task-related information. For instance, an optimal decision-maker would take into account the distribution of the opponent's actions and, more simply, the payoffs in order to gauge whether exploration is likely to yield a better strategy than the best known one. A cognitive model of human decision-making in such situations would describe the cognitive process underlying task execution and include well-known framing effects [29].

Given a theory of how individual predispositions interact with task familiarity and other meta-cognitive insights, we will be able to suggest ways that, e.g., discourage risk-taking by individuals who fail to analyze tasks explicitly, or reduce policy-based restrictions in cases where an organization could benefit from increased freedom. While such utopian ideas may still be far off, consequences are abound in measures such as selective surveillance. Further, without studies about fundamental factors of human security-related behavior, we cannot set a path towards such visionary interventions.

VI. Conclusion

Differences in individuals' cognitive predispositions lead to significant and non-obvious biases during the timing of security decisions. Our experiments apply to a range, but not all of the decisions made in computer security. We focus on rapid choices made by individual human decision-makers in real-time, rather than those made as a matter of policy-setting.

Naturally, the risk/reward structure chosen may affect decision-makers, as well as the framing of the task. (Our experiment was not framed).

Provided these caveats are understood, some of the results may surprise the system designer. Individuals that prefer to make thoughtful, deliberate decisions generally fare better over the range of task experience we studied. Other individuals, i.e., those that prefer intuitive decision-making, seem to benefit from clear risk-avoidance or even risk-seeking. The experiments show that individuals of a range of risk propensities can make successful timing decisions in principle, but not necessarily in the same kind of tasks. As could be expected, tasks with enough information available to reason carefully are suited to deliberate thinking – but in particular if risk-preferences are average. Tasks requiring more working memory and perhaps intuitive decision-making may not benefit from average risk preferences. Levels of risk propensity that maximize the outcome for some tasks can be thought of as those that are risk-neutral, maximizing utility in the security context. Interestingly, however, these "useful" amounts of risk-taking happen to be near the population average.

Our results are compatible with a theoretical view that posits the following. Cognitive predispositions vary between individuals, but some of them interact in ways that suggest that there are more or less fortunate combinations when it comes to decision-making. This appears to be the case for risk propensity and need for cognition. We hypothesize that some combinations of traits can even increase predictability.

The second, perhaps remarkable effect we observe is that risk-seeking individuals benefit more from familiarity with the task. Thus, risk propensity may be thought of as a behavior that facilitates learning. A similar picture emerges for need-for-cognition in the more difficult task of Experiment 2, where figuring out the optimal strategy takes longer. There, participants with a high need for cognition perform better once they are experienced.

Is there a trajectory for training decision-makers? There may be. Yet, our experiment naturally did not control cognitive traits and their interactions. Thus, we cannot draw conclusions about underlying mechanisms, according to which traits cause changes in outcomes. If we were, however, to select decision-makers out of a population, we would prefer moderate ones that are neither too conservative, nor too risk-seeking, and ones that tend to *think things through* for novel tasks. Assuming more experience, however, higher risk propensity would be beneficial.

The attendant question from a cognitive science perspective – and one to explore next – is how the cognitive predispositions actually combine in people. That is, do they typically occur in advantageous ways? The second question we are exploring is whether we can develop cognitive models of decision-making in security that incorporate such preferences.

The results presented in this paper illustrate an important lesson for security system design and policy: individual differences bias decision-making in predictable ways. From a cyber-security perspective, we ask whether security managers can utilize data about cognitive predispositions and begin adapting policies to individual users, or begin addressing the observed biases through intervention strategies.

VII. Acknowledgments

We appreciate the detailed feedback from the anonymous reviewers and from Lujo Bauer. The system for conducting the experiment was originally conceived by Ethan Heilman (https://github.com/EthanHeilman/flipIt) and was heavily modified and extended for the purposes of this paper by Alan Nochenson. This research is partially supported by an award from the Social Science Research Institute at the Pennsylvania State University.

References

[1] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3(1):26–33, January–February 2005.

[2] Alessandro Acquisti and Jens Grossklags. What can behavioral economics teach us about privacy? In A. Acquisti, S. Gritzalis, S. Di Vimercati, and C. Lambrinoudakis, editors, *Digital Privacy: Theory, Technologies, and Practices*, pages 363–380. Auerbach Publications, Boca Raton, FL, 2007.

[3] Alessandro Acquisti and Jens Grossklags. An online survey experiment on ambiguity and privacy. *Communications & Strategies*, 88(4):19–39, December 2012.

[4] Antoine Bechara, Antonio Damasio, Hanna Damasio, and Steven Anderson. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1–3):7–15, April-June 1994.

[5] David Blackwell. The noisy duel, one bullet each, arbitrary accuracy. Technical report, The RAND Corporation, D-442, 1949.

[6] Rainer Böhme and Jens Grossklags. The security cost of cheap user interaction. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, pages 67–82, Marin County, CA, 2011.

[7] Markus Brunnermeier and John Morgan. Clock games: Theory and experiments. *Games and Economic Behavior*, 68(2):532 – 550, March 2010.

[8] John Cacioppo and Richard Petty. The need for cognition. *Journal of Personality and Social Psychology*, 42(1):116–131, January 1982.

[9] Chris Callison-Burch. Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Singapore, August 2009.

[10] Nicolas Christin, Serge Egelman, Timothy Vidas, and Jens Grossklags. It's all about the benjamins: An empirical study on incentivizing users to ignore security advice. In *Proceedings of the Fifteenth International Conference on Financial Cryptography and Data Security (FC)*, pages 16–30, Gros Islet, St. Lucia, February - March 2011.

[11] Janice Dole and Gale Sinatra. Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, 33(2-3):109–128, Spring-Summer 1998.

[12] Eric Friedman, Mikhael Shor, Scott Shenker, and Barry Sopher. An experiment on learning with limited information: Nonconvergence, experimentation cascades, and the advantage of being slow. *Games and Economic Behavior*, 47(2):325–352, May 2004.

[13] Andrea Gallice. Preempting versus postponing: The stealing game. Technical report, ICER Working Papers 02-2008, June 2008.

[14] Moojan Ghafurian and David Reitter. Impatience, risk propensity and rationality in timing games. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society (CogSci)*, Quebec City, Canada, July 2014.

[15] Edward Glaeser, David Laibson, José Scheinkman, and Christine Soutter. Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846, August 2000.

[16] Nathaniel Good, Jens Grossklags, David Thaw, Aaron Perzanowski, Deirdre Mulligan, and Joseph Konstan. User choices and regret: Understanding users' decision process about consensually acquired spyware. *I/S: A Journal of Law and Policy for the Information Society*, 2(2):283–344, Spring-Summer 2006.

[17] Joseph Goodman, Cynthia Cryder, and Amar Cheema. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224, July 2013.

[18] J. Grossklags, S. Radosavac, A. Cárdenas, and J. Chuang. Nudge: Intermediaries' role in interdependent network security. In *Proceedings of the Third International Conference on Trust and Trustworthy Computing (TRUST 2010)*, pages 323–336, Berlin, Germany, June 2010.

[19] Jens Grossklags. Experimental economics and experimental computer science: A survey. In *Proceedings of the Workshop on Experimental Computer Science (ExpCS'07)*, San Diego, CA, June 2007.

[20] Jens Grossklags, Nicolas Christin, and John Chuang. Predicted and observed user behavior in the weakest-link security game. In *Proceedings of the 2008 USENIX Workshop on Usability, Psychology, and Security (UPSEC'08)*, San Francisco, CA, April 2008.

[21] Jens Grossklags, Nicolas Christin, and John Chuang. Secure or insure? A game-theoretic analysis of information security games. In *Proceedings of the 17th International World Wide Web Conference (WWW)*, pages 209–218, Beijing, China, April 2008.

[22] Jens Grossklags and Carsten Schmidt. Software agents and market (in)efficiency - A human trader experiment. *IEEE Transactions on System, Man, and Cybernetics: Part C*, 36(1):56–67, January 2006.

[23] Ken Hendricks, Andrew Weiss, and Charles Wilson. The war of attrition in continuous time with complete information. *International Economic Review*, 29(4):663–680, November 1988.

[24] John Horton, David Rand, and Richard Zeckhauser. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425, September 2011.

[25] Benjamin Johnson, Rainer Böhme, and Jens Grossklags. Security games with market insurance. In *Proceedings of the Second Conference on Decision and Game Theory for Security (GameSec)*, pages 117–130, College Park, MD, November 2011.

[26] James Kahan and Amnon Rapoport. Decisions of timing in bipolarized conflict situations with complete information. *Acta Psychologica*, 38(3):183–203, June 1974.

[27] James Kahan and Amnon Rapoport. Decisions of timing in conflict situations of unequal power between opponents. *Journal of Conflict Resolution*, 19(2):250–270, June 1975.

[28] Daniel Kahneman. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9):697, September 2003.

[29] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 47(2):263–291, March 1979.

[30] Aron Laszka, Mark Felegyhazi, and Levente Buttyán. A survey of interdependent security games. Technical Report CRYSYS-TR-2012-11-15, CrySyS Lab, Budapest University of Technology and Economics, Nov 2012.

[31] Aron Laszka, Gabor Horvath, Mark Felegyhazi, and Levente Buttyan. FlipThem: Modeling targeted attacks with FlipIt for multiple resources. Technical report, Budapest University of Technology and Economics, 2013.

[32] Aron Laszka, Benjamin Johnson, and Jens Grossklags. Mitigating covert compromises: A game-theoretic model of targeted and non-targeted covert attacks. In *Proceedings of the 9th Conference on Web and Internet Economics (WINE)*, pages 319–332, Cambridge, MA, December 2013.

[33] Aron Laszka, Benjamin Johnson, and Jens Grossklags. Mitigation of targeted and non-targeted covert attacks as a timing game. In *Proceedings of the Fourth Conference on Decision and Game Theory for Security (GameSec)*, pages 175–191, Fort Worth, TX, November 2013.

[34] Fred Leland. Critical decision making under pressure. *The Homeland Security Review*, 3(1):43–72, Winter 2009.

[35] Ree Meertens and Rene Lion. Measuring an individual's tendency to take risks: The risk propensity scale. *Journal of Applied Social Psychology*, 38(6):1506–1520, June 2008.

[36] Katherine Milkman, Dolly Chugh, and Max Bazerman. How can decision making be improved? *Perspectives on Psychological Science*, 4(4):379–383, July 2009.

[37] Alan Nochenson and Jens Grossklags. A Behavioral Investigation of the FlipIt Game. In *12th Workshop on the Economics of Information Security (WEIS)*, Washington, D.C., June 2013.

[38] Alan Nochenson and Jens Grossklags. An online experiment on consumers' susceptibility to fall for post-transaction marketing scams. In *Proceedings of the 22nd European Conference on Information Systems (ECIS)*, Tel-Aviv, Israel, June 2014.

[39] Richard Petty, Pablo Brinol, Chris Loersch, and Michael McCaslin. Chapter 21. The need for cognition. In Mark Leary and Rick Hoyle, editors, *Handbook of Individual Differences in Social Behavior*, pages 318–329. The Guildford Press, New York/London, 2009.

[40] Viet Pham and Carlos Cid. Are we compromised? Modelling security assessment games. In *Proceedings of the Third Conference on Decision and Game Theory for Security (GameSec)*, pages 234–247, Budapest, Hungary, November 2012.

[41] Tadeusz Radzik. Results and problems in games of timing. *Lecture Notes-Monograph Series, Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, 30:269–292, 1996.

[42] Jennifer Reinganum. On the diffusion of new technology: A game theoretic approach. *The Review of Economic Studies*, 48(3):395–405, July 1981.

[43] David Reitter, Jens Grossklags, and Alan Nochenson. Risk-seeking in a continuous game of timing. In R. West and T. Stewart, editors, *Proceedings of the 12th International Conference on Cognitive Modelling (ICCM)*, pages 397–403, 2013.

[44] Aric Shafran. Interdependent security experiments. *Economics Bulletin*, 30(3):1950–1962, July 2010.

[45] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Cranor, and Julie Downs. Who falls for Phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI)*, pages 373–382, Atlanta, GA, April 2010.

[46] Keith Stanovich and Richard West. Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23(5):645–665, October 2000.

[47] Marten van Dijk, Ari Juels, Alina Oprea, and Ronald Rivest. FlipIt: The game of "stealthy takeover". *Journal of Cryptology*, 26(4):655–713, October 2013.

[48] Na Wang, Jens Grossklags, and Heng Xu. An online experiment of privacy authorization dialogues for social applications. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, pages 261–272, San Antonio, TX, February 2013.

[49] Stacy Wood and Joffre Swait. Psychological indicators of innovation adoption: Cross-classification based need for cognition and need for change. *Journal of Consumer Psychology*, 12(1):1–13, January 2002.

[50] Ronghui Xu. Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22):3527–3541, November 2003.

Appendix

The Appendix includes the instructions for Experiment 1. The appearance of the instructions has been edited to fit the format of the proceedings.

A. Basic Rules

You will be playing multiple rounds of a two-player game called FlipIt. The objective of FlipIt is to gain and maintain possession of the game board. Until you take an action, the state of possession of the game board is hidden from your view. In this state, the board is shown in gray color.

The only action you have available is to 'flip' the game board. When you flip the board, it will be shown to you who had possession of the game board until this very moment. This information will only be shown to you and not your opponent. At the same time, you also gain possession of the board, or maintain possession if you already owned the board.

The same rules apply to your opponent. That is, you cannot observe if and when the opponent flipped the board in the past, until you take the action to flip the board yourself. Below, we break down the rules in more detail.

B. *Detailed Rules*

1) Points:

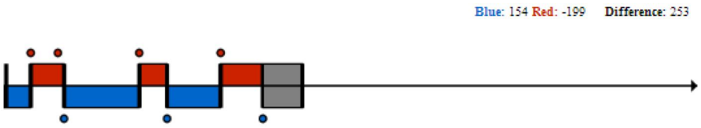
- You gain 100 points per second that you are in control.
- You earn 0 points while your opponent is in control.
- You pay 100 points when you play 'flip'.
- You begin with 0 points. Scores are updated when you play a 'flip' and at the end of the game.

2) Moves: Your only move is to play 'flip'. If you are in control and you play 'flip' you remain in control. If you are not in control and you play 'flip' you regain control. Only one player can be in control at a time.

3) The Board: The board displays the current known information about the game, including your points, the points of the red player, and the difference between your points and the points of the red player. Each 'flip' played is marked with a dot. You can only see information that was revealed by your flips. Blue rectangles represent periods of time in which you, the blue player, had control. Red rectangles represent periods of time in which the red player was in control.

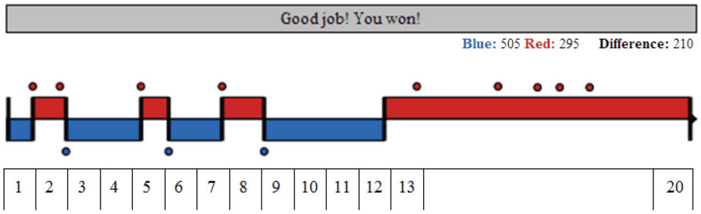
C. *An Example Game*

Fig. 7: The game in progress.



Blue: 154 Red: -199 Difference: 253

Fig. 8: The game when finished.



Good job! You won!

Blue: 505 Red: 295 Difference: 210

Let us examine the moves made in the game given above (see Figures 7 and 8.)

- *1st second*: The blue player starts in control.
- *2nd second*: The red player plays 'flip' and gains control. The red player plays 'flip' again less than a second later and remains in control.
- *3rd second*: The blue player plays 'flip' and regains control. He maintains control for a bit over 2 seconds.
- *5th second*: The red player plays 'flip' and regains control. He keeps control for less than a second.
- *6th second*: The blue player plays 'flip' and regains control. He keeps control for about 2 seconds.
- *7th second*: The red player plays 'flip' and regains control. He keeps control for about 1 second.
- *9th second*: The blue player plays 'flip' and regains control. He maintains control for 4 seconds.
- *12th second*: The red player plays 'flip' and regains control. He maintains control for the rest of the game. He makes a number of flips in which he maintains control.
- *20th second*: The game ends.

The blue player was in control for 8.05 seconds earning 805 points, and played 'flip' 3 times, costing 300 points. This gives him a total score of 505 points.

The red player was in control for 11.95 seconds earning 1195 points, and played 'flip' 9 times, costing 900 points. This gives him a total score of 295 points.

The blue player has more points than the red player and thus wins.

D. *Payment*

You will be compensated according to your performance in this study. For completing the study, you are guaranteed the amount listed on the Mechanical Turk HIT that you have accepted, and you will be paid an additional sum based on your performance. You will participate in multiple rounds of the game. At first, you will participate in a practice round without a bonus payment to familiarize yourself with the interface. Then, you will participate in several additional rounds. You can receive a bonus payment for your performance in each of those rounds.

You can increase your bonus payment in a given round by performing well compared to the red player. If you lose by more than 1000 points, however, you will receive no bonus payment for that round.

The exchange rate for points into the bonus payment is 1 cent for 100 points. For example, you would earn a bonus payment of 10 cents by gaining exactly as many points as the red player. If you outperform your opponent by 500 points you would earn 15 cents. If you underperform you opponent by 500 points you would earn 5 cents.

[Participants could continue to the game by clicking a button with the following text: "I understand the rules."]