




Network defense and behavioral biases: an experimental study

Daniel Woods¹ · Mustafa Abdallah² · Saurabh Bagchi² · Shreyas Sundaram² · Timothy Cason¹ 

Received: 28 January 2020 / Revised: 9 February 2021 / Accepted: 5 April 2021
© Economic Science Association 2021

Abstract

How do people distribute defenses over a directed network attack graph, where they must defend a critical node? This question is of interest to computer scientists, information technology and security professionals. Decision-makers are often subject to behavioral biases that cause them to make sub-optimal defense decisions, which can prove especially costly if the critical node is an essential infrastructure. We posit that non-linear probability weighting is one bias that may lead to sub-optimal decision-making in this environment, and provide an experimental test. We find support for this conjecture, and also identify other empirically important forms of biases such as naive diversification and preferences over the spatial timing of the revelation of an overall successful defense. The latter preference is related to the concept of anticipatory feelings induced by the timing of the resolution of uncertainty.

Keywords Laboratory experiment · Probability weighting · Naive diversification · Network security

This research was supported by grant CNS-1718637 from the National Science Foundation. We thank the editor, two anonymous referees, and participants at the Economic Science Association and Jordan-Wabash conferences for valuable comments.

✉ Timothy Cason
cason@purdue.edu

¹ Economics Department in the Krannert School of Management at Purdue University, Purdue University, West Lafayette, IN, USA

² School of Electrical and Computer Engineering at Purdue University, Purdue University, West Lafayette, IN, USA

1 Introduction

Economic resources spent on securing critical infrastructure from malicious actors are substantial and increasing, with worldwide expenditure estimated to exceed \$124 billion in 2019 (Gartner 2018). Cybersecurity defense is becoming increasingly difficult, as systems are frequently connected to the outside world through the Internet, and attackers innovate many new methods of attack. The interaction of computers, networks, and physical processes (termed ‘Cyber-Physical Systems’, or CPS) has a wide variety of applications, such as manufacturing, transportation, medical care, power generation and water management (Lee, 2015), and has both practical and theoretical importance. Proposed CPS such as the ‘Internet of Things’ promise vast benefits and efficiencies, but at the cost of increased attack vectors and targets (see Alaba et al., 2017; Humayed et al., 2017 for surveys). To realize the potential gains that these new technologies can provide, we must understand and maximize their security.

To reduce interference with their systems, institutions allocate a security budget and hire managers responsible for minimizing the probability of successful attacks on important assets and other vital parts of the infrastructure. Such decision-makers, however, are subject to behavioral biases that can lead to sub-optimal security decisions (Abdallah et al., 2019a, b; Acquisti & Grossklags, 2007). Human decision-makers can exhibit many possible biases. The security decisions they face broadly involve probabilistic assessments across multiple assets and attack vectors, many featuring low individual likelihood. We therefore ex-ante focus on the possibility that people incorrectly weight the actual probability of attack and defense (Tversky & Kahneman, 1992). We ex-post find that people also exhibit locational and spreading biases in their defense resource allocations, due to the directional and compartmentalized nature of these systems. Given the immense size of global expenditures on cybersecurity, as well as successful attacks being potentially very damaging, it is important to understand the nature and magnitude of any biases that can lead to sub-optimal security decisions. Such insights on biases can then be applied by security professionals to reduce their impact.

We focus on human biases as infrastructure security decisions have not yet been given over to algorithmic tools. They are still mostly made by human security managers (Paté-Cornell et al., 2018). Adoption of automated tools are stymied by legacy components in these interconnected systems, so instead managers use threat assessment tools which return the likely probability that individual components of the infrastructure will be breached (Jauhar et al., 2015). These probabilities must be interpreted by the human manager, which motivates our initial emphasis on non-linear probability weighting. Evidence also exists that security experts ignore more accurate algorithmic advice when available and instead rely more on their own expertise (Logg et al., 2019).

We model a security manager’s problem as allocating his budget over edges in a directed attack graph with the nodes representing various subsystems or components of the overall CPS. An example of a directed attack graph is shown in

Fig. 1. The manager's goal is to prevent an attacker who starts at the red node on the left from reaching the critical green node on the far right. The inter-connectivity of different systems is represented by connections between nodes, and alternative paths to a given node represent different methods of attack. Allocating more of the security budget to a given edge increases the probability that an attack through that edge will be stopped. Such an 'interdependency attack graph' model is considered an appropriate abstraction of the decision environment a security professional faces in large-scale networked systems.¹ The probability of successful defense along an edge is weighted according to the manager's probability weighting function. We use the common Prelec (1998) probability weighting function, but similar comparative statics can be obtained with any 'inverse S-shaped' weighting function. We assume the attacker is sophisticated and observes the manager's allocation decision, and does not mis-weight probabilities. This reflects a 'worst-case' approach to security (discussed further in Sect. 2.1), and represents a necessary first step in investigating the impact of probability weighting and other biases on security expenditures.

The manager's mis-weighting of probabilities can cause investment decisions to substantially diverge from optimal decisions based on objectively correct true probabilities, depending on network structure and the security production function. The security production function maps defense resources allocated to an edge to the probability that an attack along that edge will be stopped. Empirical evidence has shown probability weighting to be relatively non-linear on the aggregate subject level (Bleichrodt & Pinto, 2000), so the impact on security decisions could be substantial. Probability weighting is also heterogeneous across individuals (Tanaka et al., 2010; Bruhin et al. 2010). Therefore, if probability weighting affects choices in this environment, individuals should exhibit heterogeneity in their sub-optimal security decisions.

We seek to address the following research questions:

Question 1: What is the effect of probability weighting on security investments over a directed network graph?

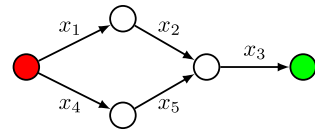
Question 2: Is probability weighting an empirically relevant factor in human security decision-making?

Question 3: What other behavioral biases significantly affect decision-making in this environment?

To address Question 1, we numerically solve the security manager's problem described above. In practical situations the relationship between investment spending and reductions in the probability of an attack is far from explicit to an outside observer. Moreover, investigations of successful breaches are often not revealed until months or years later. Furthermore, information on security investments is highly confidential for obvious reasons, making it difficult or impossible to obtain

¹ A non-exhaustive list of research considering the attack graph model from the Computer Security literature includes Sheyner and Wing (2003), Nguyen et al. (2010), Xie et al. (2010), Homer et al. (2013), and Hota et al. (2018). The length of this list and the ease in which it could be extended is indicative of the prominence that this literature places on the attack graph model.

Fig. 1 Example directed network attack graph



directly from firms. We therefore conduct an incentivized laboratory experiment to address Questions 2 and 3. We employ networks that cleanly identify the impact of non-linear probability weighting on security investment decisions, and the generated data also reveal other behavioral biases that exist in this environment.

Our experiment elicits separate measures of probability weighting outside the network defense problem to help address Question 2. One measure uses binary choices between lotteries which is relatively standard, and elicits probability weighting while controlling for the confound of utility curvature. The other measure is novel, and uses a similar network path framing to the network defense environment. This new measure reduces procedural variance relative to the main network defense task. It also exploits the irrelevance of utility curvature when there are only two outcomes to focus solely on probability weighting.

We find that the network-framed measure of non-linear probability weighting is statistically significantly correlated with all the network defense allocations situations we consider. However, this correlation exists even in cases where probability weighting should have no impact. This suggests that subjects may exhibit limited sophistication beyond probability weighting alone. We therefore conduct a cluster analysis to identify heterogeneous patterns of behavior not predicted by probability weighting. This identifies additional behavioral biases. The first is a form of ‘naive diversification’ (Benartzi & Thaler, 2001), where subjects have a tendency towards allocating their security budget evenly across the edges. The second is a preference for stopping the attacker earlier or later along the attack path. Stopping an attack earlier can be seen as reducing the anticipatory emotion of ‘dread’ (Loewenstein, 1987) while stopping it later can be seen as delaying the revelation of potentially bad news (e.g., see Caplin & Leahy, 2004 for a strategic environment). Accounting for these additional biases, we continue to find some evidence that non-linear probability weighting influences subject behavior, as well as strong evidence for the additional biases. In our environment the additional biases seem especially naive, as edges are not different options with benefits beyond defending the critical node, and information on the attacker’s progress is not presented to the subjects sequentially. These inconsistencies possibly reflect a subject’s own mental model (e.g., of how an attack proceeds), but should be accounted for in future directed network decision environments.

This paper contributes to the theoretical literature on attack and defense games over networks of targets, most of which can be related to computer network security in some fashion.² Our attack graph environment is rather flexible, and can represent

² A non-exhaustive list of related theory papers include Clark and Konrad (2007), Acemoglu et al. (2016), Dziubiński and Goyal (2013), Goyal and Vigier (2014), Dziubiński and Goyal (2017), Kovenock and Roberson (2018), and Bloch et al. (2020).

some of the strategic tensions present in alternative network environments. Instead of focusing on attack graph representations of these other environments (which can be quite complex), we utilize more parsimonious networks in order to specifically parse out the effect of probability weighting. We have the ‘security manager’ play against a sophisticated computerized attacker who moves after observing the manager’s allocation. Playing against a computer dampens socially related behavioral preferences.³ It also removes the need for defenders to form beliefs about the attacker’s probability weighting. This allows us to more cleanly identify the empirical relevance of non-linear probability weighting in this spatial network defense environment. If probability weighting is important empirically, then future research should incorporate it into models to better understand the decisions of real-world decision-makers.

This paper also contributes to the experimental literature of attack and defense games in network environments.⁴ One set of related experimental studies test ‘Network Disruption’ environments. McBride and Hewitt (2013) consider a problem where an attacker must select a node to remove from a partially obscured network, with the goal to remove as many edges as possible. Djawadi et al. (2019) consider an environment where the defender must both design the network structure as well as allocate defenses to nodes, with the goal of maintaining a network where all nodes are linked after an attack. Hoyer and Rosenkranz (2018) consider a similar but decentralized problem where each node is represented by a different player. Our environment differs from these Network Disruption games as we consider a directed attack graph network, i.e. the attacker must pass through the network to reach the critical node rather than remove a node to disrupt the network. Some other related experimental papers include ‘multi-battlefield’ attack and defense games, such as Deck and Sheremeta (2012), Chowdhury et al. (2013) and Kovenock et al. (2019). The most closely related of these types of papers is Chowdhury et al. (2016), who find experimental evidence for the bias of salience in a multi-battlefield contest, which induces sub-optimal allocations across battlefields. We are the first to investigate empirically the bias of probability weighting in networks and attack and defense games.

³ Sheremeta (2019) posits that things such as inequality aversion, spite, regret aversion, guilt aversion, loss aversion (see also Chowdhury, 2019), overconfidence and other emotional responses could all be important factors in (non-networked) attack and defense games. Preferences and biases have not received substantial attention in the experimental or theoretical literature in these games, although it should be noted that Chowdhury et al. (2013) and Kovenock et al. (2019) both find that utility curvature does not appear to be an important factor in multi-target attack and defense games.

⁴ See Kosfeld (2004) for a survey of network experiments more generally.

2 Theory and hypotheses

2.1 Attacker model

In order to describe the security manager's (henceforth defender) problem, it is necessary to describe and justify the assumptions we make about the nature of the attacker that he faces. As our focus is on network defense by humans, in our main experimental task we automate the role of the attacker and describe their decision process to a human defender. We assume that the attacker observes the defender's decision, has some fixed capability of attack, and linearly weights probabilities. While these assumptions may seem strong, they are consistent with a 'worst-case' approach, the motivation of which we now describe.

Due to the increasing inter-connectivity of cyber-physical systems to the outside world (e.g. through the internet), a defender faces a wide variety of possible attackers who can differ substantially in their resources, abilities and methods. The defender could undertake the challenging exercise of considering the attributes of all possible attackers, but this would involve many assumptions that the defender might get wrong. Instead, we assume that the defender takes a worst-case approach and defends against a sophisticated attacker, so that he can achieve a certain level of defense regardless of what type of attacker eventuates. The sophisticated attacker can be interpreted as the aggregate of all attackers perfectly colluding. They may also have the ability to observe the defender's decision either through a period of monitoring or by using informants. Taking a worst-case approach is common in the security resource allocation literature (e.g. Yang et al. 2011; Nikoofal and Zhuang 2012; Fielder et al. 2014), as is the assumption that the attacker observes the defender's allocation.⁵

2.2 Defender model

The defender faces a network consisting of J total paths from the start node to the critical node, with each edge belonging to one or more of the J paths. The defender's security decision is to allocate a security budget of $B \in \mathbb{R}_{>0}$ units across the edges; this is represented by a vector x with N elements, where N is the number of edges. The edge defense function $p(x_i)$ is a production technology that transforms the number of units allocated to edge i (denoted by x_i) to the probability of stopping an attack (from the worst-case attacker) as it passes along edge i . We assume the defender has probability weighting from the one parameter model described in Prelec (1998), i.e. $w(p(x_i); \alpha) = \exp[-(-\log(p(x_i)))^\alpha]$ with $\alpha \in (0, 1]$, although our findings hold with other 'inverse-S' shaped weighting functions (e.g., Tversky and Kahneman 1992). For ease of notation we will frequently shorten $w(p(x_i); \alpha)$ to $w(p)$ or $w(p(x_i))$.

⁵ For example, Bier et al. (2007), Modelo-Howard et al. (2008), Dighe et al. (2009), An et al. (2013), Hota et al. (2016), Nithyanand et al. (2016), Guan et al. (2017), Wu et al. (2018), and Leibowitz et al. (2019).

The defender gains a payoff of 1 if the critical node is not breached by the attacker, and gains a payoff of 0 if the attacker breaches the critical node. As the attacker observes the defender's allocation and chooses the objectively most vulnerable path (i.e. the attacker has $\alpha = 1$), the attacker's action directly follows from a given allocation. However, the defender's non-linear weighting of probabilities ($\alpha < 1$) may cause him to have a different perception about which paths are the most vulnerable. Thus, the defender thinks the attacker will choose the path with the lowest *perceived* probability of successful defense (from the defender's perspective, in accordance with their probability weighting parameter). The defender's goal is to maximize his perceived probability of successfully defending the critical node, which is determined by his weakest perceived path. The defender's optimization problem depends on the network structure, edge allocations, edge defense function $p(x_i)$, and his probability weighting parameter α . We denote the defender's overall perceived probability of defense along path j as $f_j(x; \alpha)$.

An attacker passing along an edge to reach a specific node is a separate and independent event from all other edges.⁶ We assume the defender applies his weighting function to each probability individually before calculating the probability of overall defense along a path. The defender ranks the event of stopping an attack along a given edge higher than the event of an attack proceeding. Therefore, in accordance with rank dependent utility (RDU) (Quiggin, 1982) and cumulative prospect theory (CPT) (Tversky & Kahneman, 1992), he applies his weighting function to the probability of stopping an attack along an edge ($w(p)$), and considers the other event (the attack proceeding) to have a probability of $1 - w(p)$. Therefore, a path j with three edges has an overall perceived probability of defense of $f_j(x; \alpha) = w(p(x_1)) + [1 - w(p(x_1))] [w(p(x_2)) + (1 - w(p(x_2)))w(p(x_3))]$.⁷ The defender's constrained objective problem is presented in Eq. (1).

$$\begin{aligned} \underset{x}{\operatorname{argmax}} \quad & \min \{f_1(x; \alpha), f_2(x; \alpha), \dots, f_J(x; \alpha)\} \\ \text{s.t.} \quad & x_i \geq 0, i = 1, 2, \dots, N \\ & \sum_{i=1}^N x_i \leq B \end{aligned} \quad (1)$$

We now consider the impact that non-linear probability weighting by a defender has on various network structures and defense production functions. We analyze the situation in a general setting, before considering the experimental design that we implement in the laboratory.

⁶ The events are independent as each edge represents a unique layer of security that is unaffected by the events in other edges/layers of security. Breaches of other layers of security can affect whether a specific layer is encountered, but they do not change the probability that layer is compromised.

⁷ This approach is similar to the concept of 'folding back' sequential prospects, as described in Epper and Fehr-Duda (2018) with regards to 'process dependence'. The alternative (i.e., $f_j(x; \alpha) = w(p(x_1)) + [1 - w(p(x_1))] [w(p(x_2)) + (1 - w(p(x_2)))w(p(x_3))]$) does not yield interesting comparative statics in α due to the monotonicity of the probability weighting function, so we do not consider it further.

2.3 Common edges

The described objective in Eq. (1) is a straightforward constrained optimization problem. Unfortunately, the problem is analytically intractable and no closed-form solution exists. Consider our first type of network structure, presented in Fig. 1. The key feature of this network is that one of the edges is common to both paths, while the other edges belong only to the top or bottom path. We denote $x_3 = \frac{y_3}{z}$ and assume that $v = x_1 = x_2 = x_4 = x_5$, an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{z}}$ (where z is some normalization parameter), and that $v > 0$, $y > 0$. Even with these simplifications and assumptions, taking the first order conditions of the associated Lagrangian yields a set of equations that is intractable to solve for a closed form solution for either y or v .⁸ Fortunately, it is possible to numerically solve the defender's optimization problem. For example (and anticipating our experimental design), when $z = 18.2$, $B = 24$ and $\alpha = 0.6$, the optimal allocation is $v = x_1 = x_2 = x_4 = x_5 = 1.26$ and $y = x_3 = 18.96$. Appendix A provides more analysis on how the numerical solution is calculated and whether the solution is unique.

The main trade-off in this type of network is the allocation to edges that are common to both paths or to edges that are only on one path. Consider taking a small amount ϵ from the common edge x_3 and placing it on a non-common edge. Placing the ϵ only on one edge is non-optimal for any α , as the sophisticated attacker will attack the weaker path, meaning ϵ should be split across paths. This need to split over paths reduces the marginal impact of units allocated to the non-common edges on the overall probability of defense, making them relatively less attractive compared to the common edge. However, with non-linear probability weighting ($\alpha < 1$), small probabilities are over-weighted, i.e. perceived to be higher than their actual probabilities. This increases the perceived impact of units placed on non-common edges, and can exceed the loss of having to split the allocation across more than one path. This makes expenditures on non-common edges more likely for those with non-linear probability weighting.

We can confirm this intuition numerically for a variety of edge defense functions. We mainly consider concave functions in our experiment, which have a natural interpretation of diminishing marginal returns of production.⁹ In particular, consider the edge defense function from before ($p(x_i) = 1 - e^{-\frac{x_i}{z}}$). Figure 2 plots the optimal amount to allocate to the common edge for different values of z and different levels of probability weighting α . At $\alpha = 1$ the optimal allocation is to place all $B = 24$ units on the common edge. A defender with $\alpha = 1$ will always place all of his units on the common edge for the exponential family of edge defense functions (Abdallah et al., 2019b). As α decreases, i.e., the defender exhibits increasing levels of non-linear probability weighting, he places fewer units on the common edge (and more units on the non-common edges).

⁸ Weighting the probability of a successful attack along an edge instead is analytically tractable as terms conveniently cancel, as shown in Abdallah et al. (2019b). However, this would be inconsistent with how events are ranked and weights are applied in RDU and CPT. Despite the lack of symmetry in the one parameter Prelec weighting function, the qualitative comparative statics presented in Abdallah et al. (2019b) have been numerically confirmed to hold in the current environment.

⁹ Concavity and diminishing marginal returns is a common assumption in the computer security literature (e.g., Pal and Golubchik 2010; Boche et al. 2011; Sun et al. 2018; Feng et al. forthcoming)

Consider next a non-exponential edge defense function $p(x_i) = (\frac{x_i}{z})^b$, where z is again a normalization factor and $b \in (0, \infty)$. If $b < 1$, this function is concave, if $b = 1$ it is linear and if $b > 1$ it is convex. Figure 3 illustrates that regardless of the convexity of the edge defense function, the amount allocated to the common edge decreases as α decreases from 1. Note also that for concave functions of this form, it is no longer optimal for $\alpha = 1$ defenders to place all of their allocation on the common edge. This is because the slope of the edge defense function for small values is sufficiently steeper than the slope of the function when all units are allocated to one edge. To see this, consider some $p(x_i)$ and denote the number of units allocated to the non-common edges as v , and the number of units allocated to the common edge as y . Denoting the overall probability of a successful defense as $F(v, y)$, then: $F(v, y) = p(\frac{v}{4}) + (1 - p(\frac{v}{4}))(p(\frac{v}{4}) + (1 - p(\frac{v}{4}))p(y))$. Taking the first order conditions: $\frac{\partial F(v, y)}{\partial v} = \frac{1}{2}p'(\frac{v}{4})[1 - p(\frac{v}{4}) - p(y) - p(\frac{v}{4})p(y)]$ and $\frac{\partial F(v, y)}{\partial y} = p'(y)[1 - 2p(\frac{v}{4}) + p(\frac{v}{4})^2]$. At the boundary solution corresponding to $v = 0$ and $y = B$, if $p(0) = 0$ the above expressions show that allocating all units to the common edge is optimal if $p'(0)(1 - p(B)) \leq 2p'(B)$, i.e., the marginal return to placing another unit on y exceeds that of v at the boundary. It follows that if the slope is sufficiently steep for small v 's (i.e. $p'(0) > \frac{2p'(B)}{1-p(B)}$), then an $\alpha = 1$ defender will allocate a strictly positive amount to non-common edges.¹⁰

These observations lead to our first testable hypotheses:

Hypothesis 1 *The amount allocated to common edges (weakly) decreases as α decreases from 1.*

Hypothesis 2 *If $p'(0) > \frac{2p'(B)}{1-p(B)}$ (such as for a concave power function), then a decision-maker with linear probability weighting ($\alpha = 1$) will allocate a strictly positive amount to non-common edges.*

We now present the three color-coded networks from our experiment that are designed to explore these two hypotheses.

2.3.1 Network Red

Network Red employs the network structure presented earlier in Fig. 1, and has an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$.¹¹ According to Hypothesis 1, a defender with $\alpha < 1$ will place less than 24 units on the common edge, and the amount placed on the

¹⁰ Any $\alpha \in (0, 1]$ defender is making a similar trade-off of $\frac{\partial F(v, y)}{\partial v}$ against $\frac{\partial F(v, y)}{\partial y}$, either equating them if the solution is interior, or allocating to whichever is greater at the boundary. We do not present these first order conditions here as they are not as succinct due to the presence of $w(p; \alpha)$, although we do report the first order condition in Appendix A. Where exactly the trade-off is resolved depends on α as well as the specific functional form of $p(x_i)$. This is why the optimal allocation differs over α for a given $p(x_i)$, as well as over different $p(x_i)$ for a given α . Both patterns are displayed in Figs. 2 and 3.

¹¹ The normalization factor $z = 18.2$ was chosen such that 1 unit allocated to an edge would yield a commonly overweighted probability ($p = 0.05$), while 24 units allocated to an edge would yield a commonly underweighted probability ($p = 0.73$).

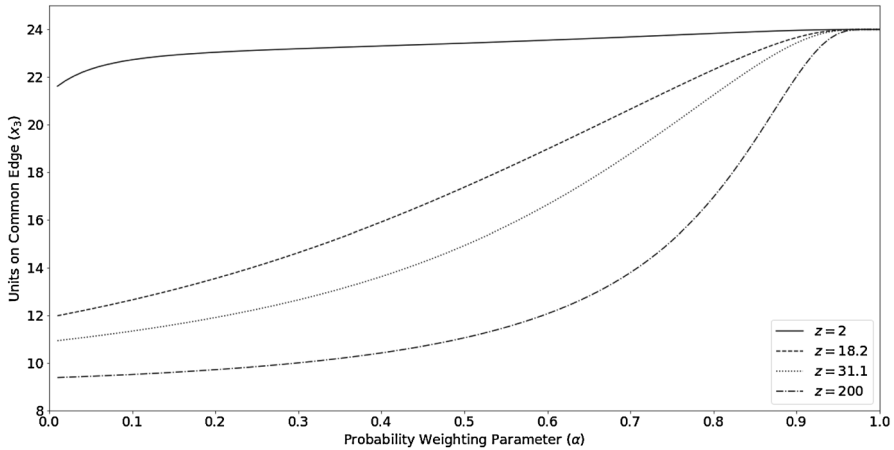


Fig. 2 Allocation to common edge for $p(x_i) = 1 - e^{-\frac{x_i}{z}}$

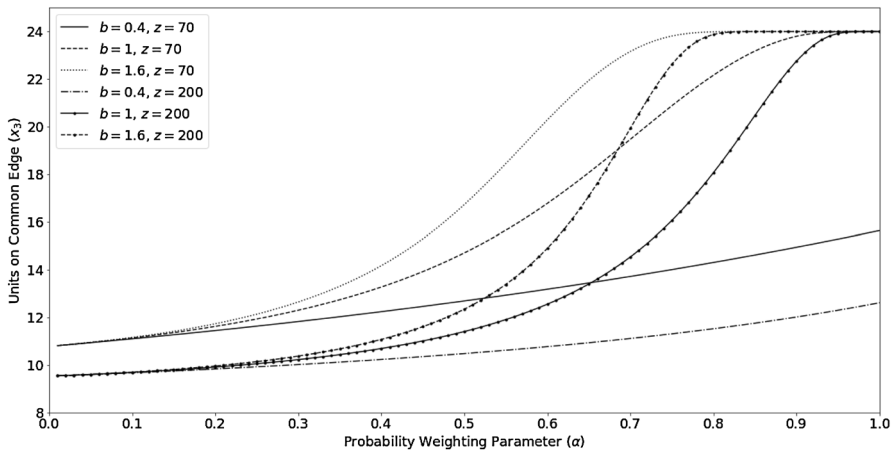


Fig. 3 Allocation to common edge for $p(x_i) = (\frac{x_i}{z})^b$

common edge is decreasing as α decreases from 1. For example, a defender with $\alpha = 0.5$ will allocate $x_3 = 17.36$, and $x_1 = x_2 = x_4 = x_5 = 1.66$, while other α 's are displayed graphically in Fig. 2 by the line associated with $z = 18.2$.¹² According to Hypothesis 2, a defender with $\alpha = 1$ would allocate $x_3 = 24$, and $x_1 = x_2 = x_4 = x_5 = 0$.

¹² These numerical solutions are continuous, although subjects were restricted to discrete (integer-valued) allocations.

2.3.2 Network Orange

Network Orange also takes place on the network shown in Fig. 1, but differs in having an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{31.1}}$. The prediction for a defender with $\alpha = 1$ remains unchanged from Network Red. Because $p(x_i) \leq 0.46 \forall x_i \in [0, 24]$, edge allocations in Network Orange mostly result in probabilities that a defender with $\alpha < 1$ will overweight. Therefore, the predictions for a defender with a particular value of $\alpha < 1$ will differ from Network Red. For example, a defender with $\alpha = 0.5$ will now allocate $x_3 = 14.92$, and $x_1 = x_2 = x_4 = x_5 = 2.27$. The prediction for other α 's is displayed in Fig. 2 on the line associated with $z = 31.1$. The change in the edge defense function increases the separation of behavior between moderate to high levels of non-linear probability weighting, increasing our ability to detect differences between α types.

2.3.3 Network Yellow

Network Yellow also takes place on the network shown in Fig. 1. The edge defense function is now of a different concave functional form, $p(x_i) = \frac{x_i^{0.4}}{70^{0.4}}$. Unlike Networks Red and Orange, it is now optimal for a non-behavioral defender to allocate units to the non-common edges, in accordance with Hypothesis 2. In particular, a defender with $\alpha = 1$ will allocate $x_3 = 15.64$, and $x_1 = x_2 = x_4 = x_5 = 2.09$, while a defender with $\alpha = 0.5$ will allocate $x_3 = 12.68$, and $x_1 = x_2 = x_4 = x_5 = 2.83$. Predictions for other α 's are presented in Fig. 3, on the line associated with $z = 70$, $b = 0.4$.

Networks Red, Orange, and Yellow are jointly designed to test Hypotheses 1 and 2. In all three of these networks, the amount allocated to the common edge should decrease as α decreases, according to Hypothesis 1. In Networks Red and Orange, Hypothesis 2 predicts that those with $\alpha = 1$ should place all 24 units on the common edge, while in Network Yellow, Hypothesis 2 predicts those with $\alpha = 1$ should place less than 24 units on the common edge.

2.4 Extraneous edges

Consider the network displayed in Fig. 4. The new feature of this network is the edge denoted x_3 , which creates a third possible path from the red node to the green node. In this network, a defender's overall perceived probability of defense is:

$$F(x; \alpha) = \min \left\{ w(p(x_1); \alpha) + [1 - w(p(x_1); \alpha)] w(p(x_2); \alpha), \right. \\ w(p(x_4); \alpha) + [1 - w(p(x_4); \alpha)] w(p(x_5); \alpha), \\ w(p(x_1); \alpha) + [1 - w(p(x_1); \alpha)] [w(p(x_3); \alpha) \\ \left. + (1 - w(p(x_3); \alpha)) w(p(x_5); \alpha)] \right\} \quad (2)$$

Call the possible paths as top (through x_1 then x_2), middle (through x_1 , then x_3 , then x_5), and bottom (through x_4 then x_5). The optimal allocation will always equalize the perceived probability of successful defense for these three paths. Otherwise, the defender could increase utility by allocating an infinitesimal amount from a non-minimum path to the minimum path. Suppose $x_1 = x_2 = x_4 = x_5 = \frac{B}{4}$. The top, middle, and bottom paths all have the same perceived probability of successful defense at this allocation. Taking an infinitesimal ϵ from any (or all) of these edges and placing it on x_3 increases the perceived probability of defense of the middle path, but at the expense at the top and/or bottom path, which would now become the minimum path.

This solution of $x_1 = x_2 = x_4 = x_5 = \frac{B}{4}$ and $x_3 = 0$ is unique for any $\alpha \in (0, 1)$ whenever the edge defense function has $p'(x_i) > 0 \forall x_i$. For $\alpha = 1$ with the exponential defense function the solution is not unique, since any combination that allocates $\frac{B}{2}$ to the top and bottom paths (which implies $x_3 = 0$) is an optimal solution.¹³ These results lead to our next testable Hypothesis:

Hypothesis 3 *The amount allocated to extraneous edges is 0, and is invariant in α .*

2.4.1 Network Blue

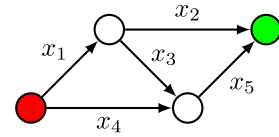
Network Blue takes place on the network with an extraneous edge, as shown in Fig. 4, with an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$. The edge defense function for Network Blue (as well as the subsequent Network Green) is the same as Network Red, which reduces the number of different edge defense functions subjects have to consider in our within-subjects design. Network Blue is designed to test Hypothesis 3, as no subject with any $\alpha \in (0, 1]$ should place any number of defense units on the extraneous edge labeled x_3 . This network is useful to identify subjects with alternative behavioral biases.

2.5 Unequal path lengths

Consider the network displayed in Fig. 5. The key feature of this network is the different number of edges on each path. With an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{z}}$, a defender with linear probability weighting should place half of his budget on each path. A defender with non-linear probability weighting, however, will place less of his budget on the top path (with more edges), and more of his budget on the bottom path (with fewer edges). To see the intuition behind this, consider a case where the defender starts with his allocation split equally across the two paths. Assume he spreads units allocated to a path equally across all edges in that path, and that the edge defense function yields a probability of less than $\frac{1}{e}$ (i.e. the Prelec inflection point) on each edge along the top path but more than $\frac{1}{e}$ on each edge along the bottom path. A defender with $\alpha < 1$ over-weights small probabilities and

¹³ Further details are presented in Appendix A.

Fig. 4 Network graph with an extraneous edge (x_3)



perceives the small investments across the many edges along the top path as providing more protection than they actually do. Conversely, the $\alpha < 1$ defender underweights large probabilities and perceives investments across the two edges along the bottom path as providing less protection than they actually do. Consequently, such a defender should reallocate his investment to equalize his perceived probability of successful defense on the two paths, and this requires shifting some of the allocation from the top path to the bottom path.

This is also true for any combination of the top and bottom path edge probabilities that are above or below the inflection point of the probability weighting function. If both are above the inflection point, the defender perceives both paths as being weaker than they actually are, but perceives the bottom path as being relatively weaker due to the more extreme under-weighting of larger probabilities. Similar logic applies if both probabilities are below the inflection point, since the over-weighting is stronger for the smaller probabilities along the top path.

Again, the analytical solution proves intractable, but Fig. 6 shows the numerical solutions considering the total relative allocations to edges in the top and bottom paths of the exponential edge defense functions of $p(x_i) = 1 - e^{-\frac{x_i}{z}}$ for various values of z . The overall optimal allocation for $\alpha \in (0, 1)$ occurs when equally spreading the total allocation to a path across each edge along a path, and this optimal allocation is unique. For example, for $\alpha = .9$ the optimal allocation is $x_1 = x_2 = x_3 = x_4 = x_5 = 2.23$, and $x_6 = x_7 = 6.43$. For $\alpha = 1$ the solution is not unique, as any solution that allocates $\frac{1}{2}$ over the top and bottom paths is an optimal allocation. These results lead to our final testable hypothesis:

Hypothesis 4 *The total amount allocated to a path with more edges decreases as α decreases from 1.*

2.5.1 Network Green

Network Green takes place on the network shown in Fig. 5, again with an edge defense function of $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$. It is designed to test Hypothesis 4, as defenders with $\alpha < 1$ should place fewer units on paths with fewer edges. For example, a defender with $\alpha = 0.5$ would place $x_1 = x_2 = x_3 = x_4 = x_5 = 0.964$ on each edge in the top path, and $x_6 = x_7 = 9.59$ on each edge in the bottom path.

The most simple network that could address Hypothesis 4 is that of 2 edges for one path and 1 edge for the other path. However, we deliberately exaggerated the difference between the two paths in Network Green by having 5 edges on the top path and 2 edges on the bottom path. This results in increased separation in predicted behavior between subjects with different α 's.

Fig. 5 Network graph with unequal path lengths

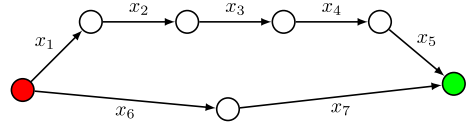


Table 1 summarizes the predictions for all five networks in the experiment for three levels of α .

3 Experimental design

3.1 Probability weighting elicitation

Our main ex ante research question as well as our hypotheses focus on how the level of non-linear probability weighting affects security investment decisions. To directly relate subjects' allocation behavior to probability weighting, we would like some external measure of probability weighting. In other words, we wish to have an accurate measure of α that has good internal validity with the network security problem, while also not taking a substantial period of time away from the main Network Defense Task.

Many ways exist to elicit an individual's probability weighting parameter. Typically researchers control for or simultaneously elicit risk preferences (taken here to mean the curvature of the utility function) when measuring probability weighting. This is because the specific range of probability weighting parameters that are consistent with a decision depends on the level of utility curvature assumed, and vice versa. However, in the defender's problem considered here, utility curvature does not play a role as there are only two payoff outcomes. The defender either successfully defends the critical node, or does not. This means that the defender always wants to maximize their (perceived) probability of the high payoff outcome, which is invariant to utility curvature. Therefore, for the Network Defense Task we are not concerned about risk preferences, other than to parse out their effect to obtain an accurate measure of probability weighting.

With that in mind, we employ a new Network *Attack* Task as a way to measure probability weighting. In this task, we have subjects swap roles, i.e., they encounter a simplified version of this network environment in the role of an attacker against a computerized defender. Not only does this elicitation task reduce the procedural variance with respect to the main defense task, it also exploits the irrelevance of utility curvature in situations with two outcomes.

Consider the network in Fig. 7, where the attacker's goal is to successfully compromise the critical node, by choosing the top or bottom path to attack. The attacker receives 3000 points for compromising the critical node, and 0 points otherwise, meaning there are only two payoff outcomes. The numbers given on each edge represent the probability of a successful attack along this edge. Because the subject plays the role of an attacker (who ranks a successful attack along an edge higher

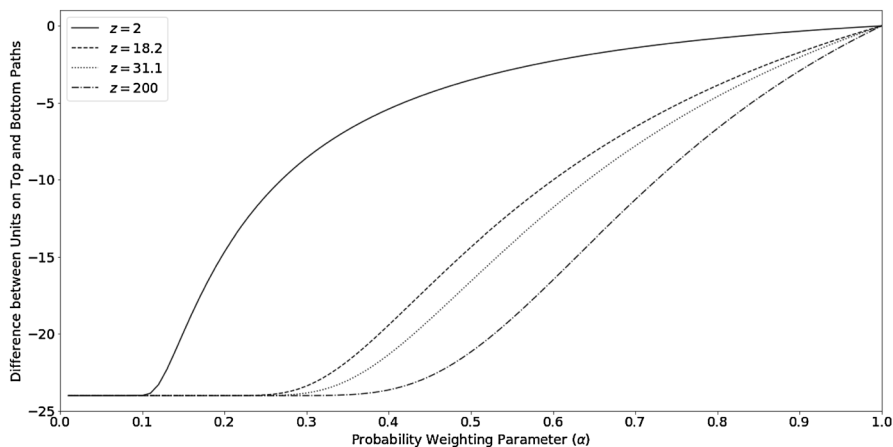


Fig. 6 Allocation to top (long) minus bottom (short) path for $p(x_i) = 1 - e^{-\frac{x_i}{z}}$

Table 1 Theoretical predictions for selected α with $B = 24$

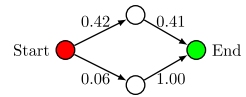
	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
Network Red $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$			
Network Orange $p(x_i) = 1 - e^{-\frac{x_i}{31.1}}$			
Network Yellow $p(x_i) = \frac{x_i^{1.4}}{10^{0.4}}$			
Network Blue $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$			
Network Green $p(x_i) = 1 - e^{-\frac{x_i}{18.2}}$			

*Any combination that allocates $\frac{B}{2}$ units to the top and bottom paths is optimal for $\alpha = 1$

than an unsuccessful attack) in this preliminary task, he weights the probability of successful attack along an edge when making his decision.

An attacker with $\alpha = 1$ should choose the top path, which has a greater probability of overall success than the bottom path ($0.42 \times 0.41 = 0.1722 > 0.06 = 0.06 \times 1.00$). However, an attacker with $\alpha < 1$ may instead prefer to attack along the bottom path, due to over-weighting 0.06 and under-weighting 0.41 and 0.42. Assuming that the attacker applies his probability weighting function to each individual probability and then calculates the probability of success of each path, then the attacker would

Fig. 7 Network attack task example



choose the top path if $\alpha > 0.597$ and the bottom path if $\alpha < 0.597$. By asking for multiple responses with different probabilities (which imply different α cutoffs), α can be bounded.

Using a dynamic bisection or staircase method could recover increasingly tight bounds on α , assuming subjects respond without error. Of course, subjects typically exhibit some level of noise in their decisions. Any mistake, especially early on in the elicitation procedure, would cause a bisection method to never be able to recover the subject's true α . A dynamic method that allows for the subject to make some errors is the dynamically optimized sequential experimentation (DOSE) method, as described in Chapman et al. (2018). In DOSE, the most informative question is asked based on the current Bayesian update of the subject's parameters. The subject's response is then used to update the current belief that the subject is of a given type, and this is then used to ask another question. The DOSE process recovers from errors as specific α types are not ruled out completely as being the subject's true type after an inconsistent response. Therefore, a subject's consistent future responses can raise the procedure's belief of the true type, and adapt future questions accordingly. DOSE always asks the most informative question given the current belief distribution over types, meaning that fewer questions are required for an accurate measure. A full description of the DOSE procedure that was implemented for this task is presented in Appendix B.

One potential concern with the Network Attack Task is that calculating the probability of a successful attack along a path is simply a case of multiplying the probabilities along the path.¹⁴ Subjects may instead perform this step before applying their subjective probability weighting, instead of after as we have assumed. We therefore take two steps to make it more difficult for a subject to trivially multiply along paths. First, we avoid using probabilities that are more easily multiplied together (such as those that end in multiples of 0 or 5). Second, we did not allow subjects to use writing utensils or calculators during this task.

In our analyses we focus on α estimates from this Network Attack Task because it reduces procedural variance from the main Network Defense Tasks and focuses solely on probability weighting. We also measured α using binary lottery choices derived from multiple price lists (MPL) used to measure probability weighting (e.g., Tanaka et al. 2010; Bruhin et al. 2010). Subjects choose between two lotteries one at a time (i.e., consider one row of an MPL in isolation), again using the DOSE procedure, also estimating an additional risk preference parameter. Details are also presented in Appendix B.

¹⁴ Another potential issue for both tasks is that subjects may not understand this point at all, instead of finding it simple. The number of such subjects should be limited due to our subject pool being drawn from a university student population.

3.2 Network defense tasks

For the Network Defense Tasks, subjects had a 24 ‘defense unit’ budget to use each period. These defense units could be allocated in integer amounts across edges. Defense units not used in one period did not roll over to the next period (i.e., this was a ‘use it or lose it’ situation). Subjects could submit a defense allocation of less than 24 units, but the software would prompt them to confirm they actually wanted to submit such an allocation.¹⁵ Subjects chose the number of defense units to allocate to an edge using a dropdown menu that automatically updated the possible options based on the remaining number of units available. The initial value of this dropdown menu was not a number, meaning subjects had to make a selection for each edge, even if the desired allocation was zero. An example of the interface is shown in Appendix F. Subjects play each of the five different networks 10 consecutive times to allow for some feedback and learning. The ordering of these five blocks was varied randomly across subjects.

3.3 Procedures

The experiments were conducted at the Vernon Smith Experimental Economics Laboratory (VSEEL). In total, 91 subjects participated, all students at Purdue University recruited from a subject database using ORSEE (Greiner, 2015).¹⁶ Subjects received a packet of written instructions, some of which were printed on color paper that aligned with the color of the Network Defense Task.¹⁷ Subjects were instructed to refer to specific instructions when the software (implemented in oTree (Chen et al., 2016)) prompted them to do so. Subjects participated in the Binary Lottery Task first, followed by the Network Attack Task. During these first two tasks, as noted above subjects were not allowed to use calculators or writing utensils, and this was strictly enforced. Subjects then completed the colored Network Defense Tasks in an order that was varied randomly and unique to each subject. Subjects could request a calculator and pen from the experimenter during the Network Defense Tasks, due to the increased computational difficulty of these tasks. To simplify probability calculations, the instructions included a table for every network indicating how allocated defense resources mapped numerically into defense likelihood for any edge.

All payoffs were denoted in experimental points, with 350 points = \$1.00. Subjects received 3000 points in a round for successfully reaching the end node in the Network Attack Task, and 1500 points for successfully preventing the computerized attacker from reaching the end node in a Network Defense Task. One round from each task was randomly selected for payment at the end of the experiment. Subjects were able to

¹⁵ In only 5 of the 4550 total decisions did subjects allocate less than all 24 units.

¹⁶ Due to an error with the software, decision times were not recorded for 4 subjects. For consistency, we present our results only considering the remaining 87 subjects. Where the inclusion of decision times is not necessary, our results do not substantially change if the dropped observations are included.

¹⁷ These instructions are available in Appendix G.

proceed through the tasks at their own pace, with most taking between 30-90 minutes (about 45 minutes on average) and earning an average of \$20.10. To ensure subjects had read the instructions carefully, before each Network Defense Task subjects were asked to report the probability of two randomly selected rows (one from 1-12, one from 13-24) of the edge defense function for that task, and were paid an additional 50 points if they answered correctly.

4 Results

We begin the results with an overview of the probability weighting (α) elicitation from the Network Attack Task. We then consider the consistency of subject behavior between and within the Network Attack Task and Network Defense Tasks, including non-parametric tests of our Hypotheses. We then present a cluster analysis to broadly summarize the heterogeneity in the strategies that subjects employ. This identifies other possible biases that subjects exhibit. Finally, we present a regression analysis on key defense allocations that controls for the identified biases and other important factors like cognitive ability and decision time.

4.1 Network attack task

The main purpose of the Network Attack Task is to obtain an estimate of an individual subject's probability weighting, parameterized by α . A useful comparison point for our results comes from Bruhin et al. (2010), who estimate a finite mixture model on certainty equivalents for lotteries elicited over many Multiple Price Lists. They find evidence for two groups, with approximately 20% of subjects exhibiting near linear probability weighting, and the remaining 80% of subjects exhibiting non-linear probability weighting.

Figure 8 presents the CDF for the subjects' elicited α 's from the Network Attack Task. Considerable heterogeneity exists in the degree of non-linear probability weighting, so our Hypotheses predict heterogeneity in the Network Defense Tasks as well. Considering the quintiles of the distribution, we have 20% of subjects with $\alpha \geq 0.95$, 20% with $0.90 \leq \alpha < 0.95$, 20% with $0.80 \leq \alpha < 0.90$, 20% with $0.64 \leq \alpha < 0.80$, and finally the bottom quintile with $\alpha < 0.64$. This suggests the presence of both relatively linear and non-linear probability weighting groups. Our results are in line with the 20% of subjects exhibiting linear weighting as in Bruhin et al. (2010), albeit with our second highest quintile being somewhat linear as well.

Result 1 *Considerable heterogeneity exists in the inferred α from the Network Attack Task. The quintile cutoff points are $\alpha = 0.64$, $\alpha = 0.80$, $\alpha = 0.90$, $\alpha = 0.95$.*

4.2 Network defense tasks

4.2.1 Summary

Figure 9 presents the CDF's of mean defense allocations for key subject decisions in each of the five Network Defense Tasks. This also indicates substantial heterogeneity in subject behavior. In the Red and Orange networks, 26.4 and 16.1% of subjects respectively allocate all units to the common edge, and this fraction of subjects decreases to 13.8% in Network Yellow. This suggests that a relatively small proportion of subjects exhibit behavior consistent with an α near 1. However, about 40% of subjects in all three of these networks allocate less defense to the common edge than can be justified even for very low levels of α , suggesting a role for additional behavioral biases. About one-third of subjects allocate no units to the extraneous edge in Network Blue, in accordance with Hypothesis 3, while 46.0% allocate more than 1 unit on average to the extraneous edge. This again suggests a role for other behavioral biases. The Network Green CDF indicates that 19.5% of subjects allocate equal amounts to the top and bottom paths, consistent with $\alpha = 1$, and 25.3% of subjects allocate less units to the top path, consistent with $\alpha < 1$. However, over half of the subjects allocate more to the top path (and many quite substantially so), which is the opposite of what Hypotheses 4 predicts for $\alpha < 1$. Overall, the CDFs provide some casual evidence in support of probability weighting playing a role in subject behavior, but that other biases appear to influence behavior as well.

4.2.2 Subject consistency and non-parametric tests

We first consider individual subject consistency between and within network defense and attack tasks. Recall that our measure of α is derived from the Network Attack Task, which we use to test our Hypotheses in the Network Defense Task. Table 2 presents the non-parametric Spearman's ρ between the elicited probability weighting (α) and decision noise (λ) from the Network Attack Task and key average subject behavior in each of the Network Defense Tasks.^{18,19} The decision noise parameter λ is estimated by the logit function, a commonly used structure in Quantal Response Equilibria (McKelvey and Palfrey 1995).²⁰ A higher λ is consistent with less noisy behavior, meaning subjects choose their payoff maximizing action more frequently. Table 2 indicates consistency within the Network Defense Tasks as correlations are strongly statistically significant in all but one of the pairwise comparisons between these tasks. We also observe

¹⁸ We consider the same analysis including the Binary Lottery Task in Appendix D. The elicited α 's of these tasks are not correlated ($\rho = 0.166$, $p = 0.117$), suggesting the procedural differences are important, or that cognitive ability may play a role.

¹⁹ Unless otherwise stated, all p-values and statistical tests are two-sided.

²⁰ $Prob(Top) = \frac{1}{1 + e^{-\lambda(U(Top) - U(Bottom))}}$ if $U(Top) \geq U(Bottom)$, $Prob(Top) = \frac{1}{1 + e^{-\lambda(U(Bottom) - U(Top))}}$ otherwise, where $U(Top)$ is the weighted then compounded probability of successful attack multiplied by the payoff from a successful attack.

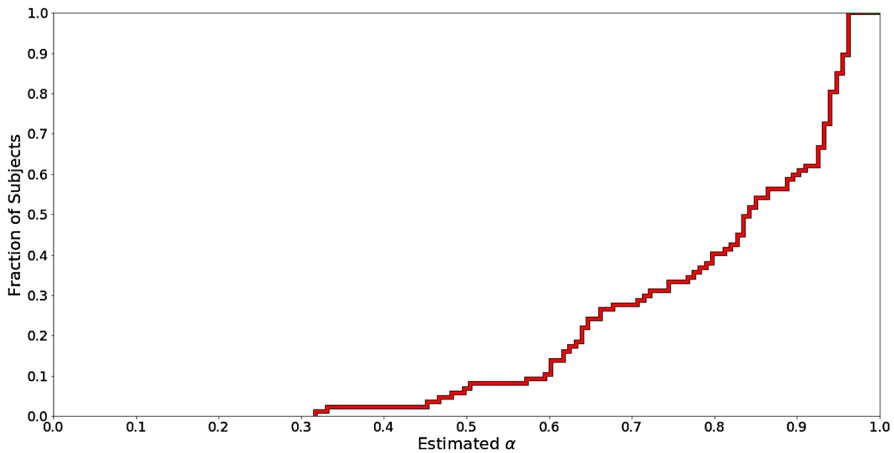


Fig. 8 CDF of elicited α from the network attack task

consistency between behavior in the Network Attack and Defense Tasks, with the leftmost column of Table 2 reporting statistically significant correlations in all but one comparison.

We now consider non-parametric tests of Hypotheses 1, 3, and 4, all of which are included in the leftmost column of Table 2. The elicited α from the Network Attack Task is strongly and significantly correlated with defense in all Network Defense Tasks except Network Yellow.²¹ The correlations of α with the common edge networks are positive, consistent with Hypothesis 1. The negative correlation in Network Green indicates that subjects with estimated α 's closer to 1 tend to place less defense on the top path. This is the opposite of Hypothesis 4. The negative correlation in Network Blue for defense resources placed on the extraneous edge provides evidence against Hypothesis 3. This suggests that our external measure of α from the Network Attack Task also captures some element of cognitive ability. This interpretation is consistent with the strong correlation of decision noise (λ) with probability weighting (α).

To conduct a non-parametric test of Hypothesis 2, we use the Wilcoxon signed-rank test. In particular, we compare the paired observations of an individual subject's average allocation to the common edge in Network Yellow to Networks Red and Orange. According to Hypothesis 2, those with an α close to 1 should exhibit a particularly pronounced decrease in this key allocation for Network Yellow. As we have a clear directional theoretical prediction we report one-sided p-values for this test. Considering subjects with an estimated $\alpha \geq .9$ (i.e. the 40th percentile closest to the linear $\alpha = 1$), we find a statistically significant decrease at the five percent level when testing Network Red against Yellow

²¹ The lack of a significant correlation in Network Yellow is not necessarily surprising, due to the deliberate reduction of the separation of α types in this network to evaluate Hypothesis 2.

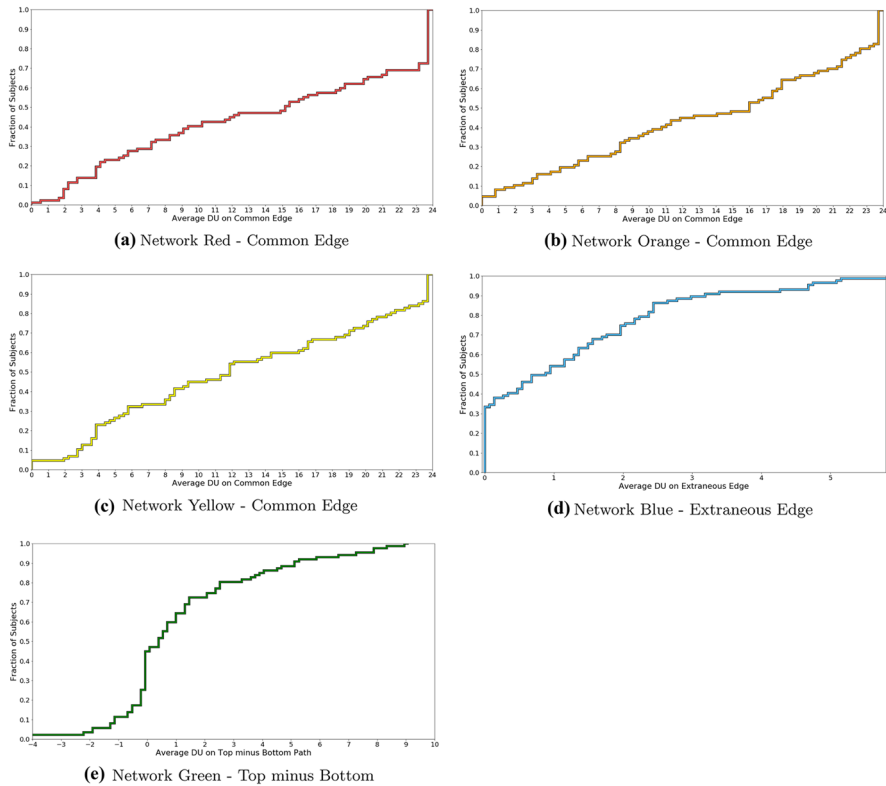


Fig. 9 CDFs of per-subject average behavior across all rounds in the network defense tasks

($p = 0.031$), as well as when we test Network Orange against Yellow ($p = 0.017$). We find no similar difference in subjects with an estimated $\alpha < .9$ ($p = 0.451$ and $p = 0.447$ respectively). These results are robust at the five percent level for any α cutoff $\in [0.86, 0.95]$ (see Appendix C).

Result 2 *The probability weighting parameter α is positively correlated with allocations to the common edge in the Red and Orange Networks, consistent with Hypothesis 1. Subjects with $\alpha \geq 0.9$ allocate less to the common edge in Network Yellow as compared to Networks Red and Orange, consistent with Hypothesis 2. α is negatively correlated with allocations to the extra edge in Network Blue and the top path in Network Green, inconsistent with Hypotheses 3 and 4.*

These initial results should be considered more as a guide to the analysis rather than an exhaustive test of our hypotheses. For instance, the Spearman correlation is a bivariate measure that does not control for any other possible biases and observed characteristics of the subject. We therefore conduct a cluster analysis in the following subsection to identify additional biases. We then conduct a

Table 2 Spearman's ρ correlation table

	α	λ	Red common edge	Orange common edge	Yellow common edge	Blue extra edge
λ	$\rho = 0.764^{***}$	$\rho = 1$				
Red common edge	$\rho = 0.260^{**}$	$\rho = 0.151$	$\rho = 1$			
Orange com- mon edge	$\rho = 0.226^{**}$	$\rho = 0.056$	$\rho = 0.654^{***}$	$\rho = 1$		
Yellow com- mon edge	$\rho = 0.072$	$\rho = -0.058$	$\rho = 0.622^{***}$	$\rho = 0.646^{***}$	$\rho = 1$	
Blue extra edge	$\rho = -0.286^{***}$	$\rho = -0.230^{**}$	$\rho = -0.352^{***}$	$\rho = -0.308^{***}$	$\rho = -0.112$	$\rho = 1$
Green top– bottom	$\rho = -0.255^{**}$	$\rho = -0.139$	$\rho = -0.241^{**}$	$\rho = -0.292^{***}$	$\rho = -0.237^{**}$	$\rho = 0.241^{**}$

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

regression analysis that controls for the identified biases and other factors like cognitive ability.

4.2.3 Cluster analysis

The previous subsection documents that probability weighting is associated with defense misallocation in this network defense environment, but not always in the manner originally hypothesized. Other behavioral biases also appear important, but these biases are not clear ex-ante. It is also likely that subjects exhibit heterogeneity in these biases and that these biases may interact, making them difficult to predict or otherwise identify. One way to summarize general patterns of subject behavior is using a cluster analysis. In particular, we use the method of Affinity Propagation (Frey & Dueck, 2007), which endogenously determines the number of clusters. We cluster at the session level, i.e., a subject's average behavior across individual networks, as we consider their behavior to be related across tasks.²²

Table 3 presents the 'exemplar' of each of the 10 clusters, summarizing an individual subject's behavior that is the most representative of that cluster. The leftmost column presents the percentage of subjects represented by that cluster, alongside a descriptive name to aid exposition.

Clusters 1 and 2 appear largely consistent with an $\alpha = 1$, meaning that approximately 17% of subjects exhibit linear probability weighting through their network defense decisions. This is close to the 20% as reported in Bruhin et al. (2010).

The cluster analysis identifies three additional biases, which along with probability weighting can describe the behavior of the cluster exemplars. The first bias is that of naive diversification: when subjects are given n options to invest in, they have a

²² We also cluster at the individual network task level in an alternative estimation presented in Appendix E. That analysis identifies similar patterns of behavior.

Table 3 Cluster analysis

Cluster Name	Network Red	Network Orange	Network Yellow	Network Blue	Network Green
C1: Near Optimal $\alpha = 1$ - Late Revelation 12.6%					
C2: Near Optimal $\alpha = 1$ - Early Revelation 4.6%					
C3: Late Revelation - Some Diversification 4.6%					
C4: Naive Diversification 16.1%					
C5: $\alpha < 1$ - Some Diversification 12.6%					
C6: Early Revelation - Some Diversification 9.2%					
C7: $\alpha < 1$ - Mild Diversification 16.1%					
C8: Naive Diversification and Late Revelation 5.7%					
C9: Late Revelation - mild Diversification 13.8%					
C10: Early Revelation - mild Diversification 4.6%					

Left column displays the percentage of subjects classified in each cluster. Numbers on edges denote average DU (out of 24 total) allocated by the exemplar subject for that cluster

tendency towards investing $1/n$ units to each option (Benartzi & Thaler, 2001). Note that this is especially *naive* naive diversification, as the edges do not represent different assets, just different ways to protect the same asset (the critical node). Naive diversification explains Cluster 4 particularly well, but can also explain situations where less units are placed on the common edge that can be justified by probability

weighting alone. A defender with $\alpha < 1$ as well as some mild preference towards evening out his allocation on the common and non common edges would place even less on the common edge than his level of α would predict. Some level of naive diversification clearly explains non-zero allocations to the extraneous edge in Network Blue. Naive diversification can also explain the tendency for subjects to place more units on the top rather than the bottom path in Network Green; as the top path has more edges, a $1/n$ heuristic would place more units overall on the top path.

The second and third biases are related to each other, and we term them early or late revelation of the overall outcome. Early revelation means that subjects try to stop the attack as soon as possible, and thus allocate more to edges nearer to the start node on the left. Clusters 6 and 10 are good examples of early revelation. Late revelation is the opposite, referring to subjects that allocate more units to edges nearer to the critical node, as exemplified by Cluster 9. Early revelation can explain an excessively low allocation to the common edge, in a manner similar to naive diversification except that more units are placed on the front two non-common edges instead of equally to all non-common edges. Late revelation can explain the failure of some subjects to reduce their common edge allocation in Network Yellow. Note that, like the naive diversification bias, this is an especially naive preference as the outcome is revealed immediately after the allocation decision is made, and importantly, all at once. In the experiment there is no animation that sequentially displays the attacker's progress. Therefore, holding anticipatory emotions such as dread over a period of time is minimized within an attack.²³ The concepts of early and late revelation are related to the literature on anticipatory utility with regards to the revelation of uncertainty (e.g., Loewenstein, 1987; Caplin & Leahy, 2001).

4.2.4 Regression analysis

The cluster analysis identifies additional biases that may interact with probability weighting and influence subject behavior. To address more directly our original hypotheses regarding the behavioral implications of probability weighting, we account for these other biases by including appropriate measures in a regression analysis. In addition to these control variables, we include additional independent variables to investigate systematically how they influence behavior.

Ex-ante we did not anticipate the additional biases, and therefore did not specifically design separate elicitation tasks to identify them. Fortunately, our Blue and Green networks allow us to measure subjects' naive diversification and early/late revelation preferences, which can then be used as controls when considering behavior in other networks.

Our main naive diversification measure is calculated from a subject's allocation to the extraneous edge in Network Blue. Specifically, we calculate each individual's average allocation to this extra edge. However, this measure clearly does not work for Network Blue, as it is based on behavior in this network. Therefore, in order to

²³ It is of course possible that subjects are playing out the attack process in their imagination, while reading the outcomes sequentially.

Table 4 Tobit regression analysis

	Red common edge	Orange common edge	Yellow common edge	Blue extra edge	Green top–bottom
α (attacker task)	15.89 (1.28)	19.06* (1.66)	27.86*** (2.73)	– 3.928 (–1.19)	– 4.341** (–2.24)
μ (attacker task)	– 0.0110 (–0.30)	– 0.0220 (–0.65)	– 0.0823*** (–2.78)	– 0.0116 (–1.12)	0.00745 (1.30)
Naive diversification [†]	– 3.395*** (–3.13)	– 2.830*** (–2.81)	– 1.713** (–2.04)	1.425*** (4.10)	0.578*** (3.40)
Early revelation [‡]	– 11.24*** (–3.65)	– 9.604*** (–3.34)	– 12.74*** (–5.32)		
Time spent on decision	– 0.0119 (–1.46)	– 0.0220** (–2.05)	– 0.00924 (–1.21)	0.00133 (0.44)	– 0.000939 (–0.32)
Total time spent on instructions	– 0.00492 (–0.57)	– 0.00598 (–0.74)	– 0.0109 (–1.61)	– 0.00604** (–2.53)	0.000438 (0.33)
Age	– 0.144 (–0.31)	0.492 (1.13)	0.615* (1.65)	0.0103 (0.08)	0.0646 (0.89)
Born in USA	– 1.107 (–0.35)	– 4.795* (–1.66)	– 2.621 (–1.09)	1.081 (1.28)	– 0.898* (–1.86)
Period	0.188** (2.18)	0.491*** (4.28)	0.0911 (1.11)	– 0.246*** (–5.15)	0.0110 (0.22)
Male	3.908 (1.25)	3.978 (1.39)	1.763 (0.73)	– 1.351 (–1.58)	– 0.200 (–0.41)
Economics major	8.190 (1.34)	3.155 (0.56)	3.147 (0.66)	0.114 (0.07)	1.167 (1.23)
Engineering major	9.388** (2.03)	7.358* (1.72)	3.273 (0.92)	1.523 (1.15)	– 1.887*** (–2.63)
Science major	2.574 (0.56)	3.318 (0.77)	1.476 (0.41)	1.443 (1.14)	– 0.500 (–0.70)
Management major	– 3.095 (–0.63)	– 0.754 (–0.17)	– 3.492 (–0.92)	– 0.0188 (–0.01)	– 1.034 (–1.34)
GPA > 3.5	2.108 (0.70)	– 6.652** (–2.38)	1.431 (0.61)	0.514 (0.62)	– 0.479 (–1.02)
Graduate student	3.447 (0.81)	– 3.169 (–0.81)	– 4.232 (–1.27)	– 1.117 (–0.94)	0.0124 (0.02)
Constant	8.911 (0.56)	– 2.588 (–0.17)	– 9.053 (–0.70)	10.54** (2.49)	3.125 (1.24)
Observations	870	870	870	870	870

t Statistics in parentheses* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ [†]Generated from Blue for Red, Orange, and Yellow, and from Green for Blue.[‡]Generated from the average of Blue and Green

obtain a measure of naive diversification for Network Blue, we use behavior from Network Green. A fully naive individual would allocate $\frac{24}{7} = 3.4$ units to each edge in Network Green, so we calculate the average absolute distance of each edge from this equal spread. A fully naive individual would have a measure of 0, and the most extreme optimal allocation of 12 units to one top and bottom edge would have a measure of $\frac{3.4 \times 5 + 8.6 \times 2}{7} = 4.88$. We then multiply this measure by -1 , so that the comparative static is comparable with the measure based on Network Blue, which has naive individuals having a higher (rather than lower) value of this measure.

For early/late revelation we consider the Blue and Green networks without the common edge, as expressing this preference is far less costly in these networks. Furthermore, early and late revelation should not impact allocations to the dependent variable in the regressions for Networks Blue and Green, so we omit this particular independent variable for those networks. The early revelation measure is based on the ratio of units allocated to the nearest two edges to the start node, and the nearest two edges to the critical node, averaged for the Blue and Green networks.

The regressions also add variables to account for cognitive ability, as the results from Table 2 suggest that α may be picking up some measure of cognitive ability.²⁴ We include information self-reported by subjects in a post-experiment survey, such as field of study, a high GPA, and whether a subject is a graduate student. We also include decision times and time spent with the instructions, as this may be correlated with subjects' understanding. Finally, we note that gender is of particular interest ex-ante based on previous observations that women tend to exhibit greater non-linear probability weighting on average than men (Fehr-Duda et al., 2006; Bruhin et al., 2010; Fehr-Duda et al., 2011).

Table 4 reports a series of censored tobit regressions, with the dependent variable for each network corresponding to the key summary statistics shown earlier in Fig. 9: the allocation to the common edge for Networks Red, Orange, and Yellow, the allocation to the extraneous edge in Network Blue, and the difference in allocations to the top and bottom paths for Network Green. The regressions are censored at 0 to 24 for all networks except Network Green, which is censored -24 to 24.

The top row shows the effect that our measure of probability weighting (α , estimated from the Attacker task) has after controlling for the identified additional biases and the other subject characteristics. Consistent with Hypothesis 1, amounts allocated to the common edge in Networks Red, Orange and Yellow are increasing in α . These coefficients are statistically significant in Networks Orange and Yellow, but not in Network Red. These results provide partial further evidence in support of Hypothesis 1 when combined with our correlation results. According to Hypothesis 3, α should not have an effect on the amount allocated to the extra edge in Network Blue. After controlling for naive diversification, we no longer find a statistically significant affect of α on this allocation. Finally,

²⁴ Choi et al. (2018) reports evidence suggesting a correlation between cognitive ability and probability weighting.

Hypothesis 4 predicts that increasing α should increase the amount allocated to the top path in Network Green. While there is a statistically significant affect of α , it is not in the direction predicted by Hypothesis 4. This is surprising because the regression controls for naive diversification, which should account for some subjects' tendency to allocate relatively more to the top path than the bottom.

Result 3 *After controlling for other biases, α is a statistically significant predictor of behavior in Networks Orange and Yellow (evidence in support of Hypothesis 1), and not in Network Blue (evidence in support of Hypothesis 3). α is a statistically significant predictor of behavior in Network Green, but in the opposite direction than predicted (evidence against Hypothesis 4).*

We now consider the impact of naive diversification, which is predicted to decrease the amount allocated to the common edge, increase the amount allocated to the extra edge in Network Blue, and increase the amount allocated to paths with more edges (i.e. the top path in Network Green). Table 4 shows that naive diversification has a negative and significant impact on the amount allocated to the common edge in all three common edge networks. Naive diversification also has a positive effect on the number of units allocated to the extra edge in Network Blue, and to the top path in Network Green, all as predicted.

Result 4 *A higher level of preference for naive diversification is correlated with a lower allocation to the common edge in Networks Red, Orange, and Yellow. It is also correlated with a higher allocation to the extra edge in Network Blue, and the longer top path in Network Green.*

Finally we consider early/late revelation, which only impacts the dependent variable for the common edge networks. Early revelation is predicted to decrease the amount allocated to the (late) common edge. The results show that a preference for early revelation has a strong and highly significant negative effect on the amount allocated to the common edge in all common edge networks.

Result 5 *A higher preference for early revelation, measured using Networks Blue and Green, is correlated with a lower allocation to the common edge in Networks Red, Orange, and Yellow.*

The other independent variables include a period variable to capture the time trend, which suggests that some learning occurs as subjects gain experience with a particular network. The only other independent variable that is statistically significant over more than two networks is whether the student was an Engineering major, which has a statistically significant effect in the direction of optimal behavior in three networks. This suggests that cognitive ability or mathematical sophistication could promote better understanding and performance in this network defense problem.

5 Conclusion and discussion

Cybersecurity and network defense is becoming increasingly important for economic, social, and even political activity. Both the financial and non-pecuniary costs of successful cyberattacks can be substantial, and thus it is important to minimize their likelihood. We investigate how behavioral biases, in particular probability weighting, could lead to sub-optimal defense allocations. We modeled the situation as a directed network graph, to capture in a simple way some trade-offs that security professionals face. Probability weighting has differing effects on various network structures and defense functions, which generates testable hypotheses. We found that a separately elicited measure of probability weighting (α) has a statistically significant correlation with key defense allocations in most Network Defense Tasks, including a network where probability weighting is predicted to have no effect. Motivated by this finding, we used a cluster analysis to identify additional biases that could also influence defense behavior. We identify preferences for naive diversification and for earlier or later revelation of attack outcomes. Controlling for these biases and other subject characteristics, we find evidence that probability weighting has predictive power in this environment, as do preferences for both naive diversification and early/late revelation.

An important question is how applicable are the findings from our student subject pool for security experts. We are not excessively concerned about this for several reasons. Firstly, a security expert may exhibit ‘other-evaluation’ (Curley et al., 1986). In the event of a successful attack, a security expert must justify his decision to others within his or her organization. If these other individuals exhibit biases, the expert may allocate in accordance to these biases to more easily justify their decision post-attack. Secondly, even if security experts exhibit fewer or weaker biases, given the magnitude of potential losses even very small biases could have a large impact on welfare. Finally, the empirical evidence on differences in behavior between students and experts is weak. Fr  chette (2015) conducts a survey of experiments that considered behavior of students compared to experts in a wide variety of professions. This survey reports only one out of thirteen considered studies found that professionals make decisions more closely in line with standard economic theory. Considering security professionals specifically, Mersinas et al. (2016) find that while security professionals do calculate expected values better than students, they also exhibit systematic biases such as ambiguity aversion and framing effects. We therefore consider the findings from our student subject pool to be sufficiently informative and useful to be taken seriously by cyber-security researchers.

Another important question is how robust our findings are to learning. We find some evidence of learning in our networks, suggesting that biases may reduce over time as subjects receive feedback and become more familiar with the task. The question is whether these biases vanish in the long run, or whether they persist. The ten rounds used for each network environment is likely insufficient for subjects to fully learn given the complexity of the environment. The number of rounds was a practical constraint, trading-off time spent in the lab against the overall number of network

structures. It would be interesting to see how behavior evolves over longer repetitions of play, but that is beyond the scope of this paper.

There are many possible avenues for future research. First, theoretical work could incorporate the additional biases into a model over directed networks. This would be a very challenging endeavor. For example, consider observing an allocation of 2 on each non-common edge and 16 on the common edge in Network Red. A wide variety of situations could be consistent with this allocation, such as $\alpha \approx 0.66$, or any $\alpha \in [.66, 1]$ with some level of naive diversification, or an $\alpha < .66$ but with some preference for late revelation, or $\alpha = 1$ having mild diversification preferences interacting with a stronger preference for late revelation, and so on. Adding to this complexity, it is not clear how the additional biases should be defined across different types of networks, or how they should interact with each other. For example, consider a subject who consistently places 2 units on the extraneous edge in Network Blue. This subject clearly has some preference for diversification, but what does that imply for his decision in Network Red? Many possibilities exist. He could be facing a minimum constraint of 2 units per edge to satisfy a diversification preference, or he could allocate 2 units to each edge initially and allocate the remaining 14 units according to his weighting parameter α (either disregarding or regarding the 2 units already allocated). Or he could be willing to give up a small amount in terms of perceived probability from his optimal strategy in order to more evenly spread his allocation, etc. Although there are many different ways that this could be modeled over different networks, the literature currently offers no guidance for explicit functional forms over directed networks to discipline these modeling decisions.

A second line of future research could incorporate strategic considerations by having human decision-makers interact with each other, in either roles of attacker and a defender, or multiple defenders on the same network defending the same or different critical nodes. For example, it may be in a defender's best interest to allocate his resources differently if he believes the attacker to have $\alpha < 1$. Alternative network structures in both the Network Defense and Network Attack Tasks could also be worth investigating, particularly in light of the identified naive diversification and early/late revelation biases. Third, it is not clear why α has a significant impact of behavior in Network Green in the opposite direction that is predicted. It may be the case that our elicitation of α is only picking up on cognitive ability. Future research could investigate why results from Network Green are anomalous, perhaps with an alternative elicitation of α or naive diversification or additional controls of cognitive ability. Finally, the effect of probability weighting in more standard attack and defense games has not yet received much attention. Given the empirical relevance of it in the current environment, this may prove to be an interesting avenue to explore.

Supplementary Information The online version supplementary material available at <https://doi.org/10.1007/s10683-021-09714-x>.

References

- Abdallah, M., Naghizadeh, P., Hota, A. R., Cason, T., Bagchi, S., & Sundaram, S. (2019). Protecting assets with heterogeneous valuations under behavioral probability weighting. In *2019 IEEE conference on decision and control (CDC)* (pp. 5374–5379).
- Abdallah, M., Naghizadeh, P., Hota, A. R., Cason, T., Bagchi, S., & Sundaram, S. (2019). The impacts of behavioral probability weighting on security investments in interdependent systems. In *2019 American control conference (ACC), Philadelphia* (pp. 5260–5265).
- Acemoglu, D., Malekian, A., & Ozdaglar, A. (2016). Network security and contagion. *Journal of Economic Theory*, 166, 536–585. <https://doi.org/10.1016/j.jet.2016.09.009>. ISSN 10957235.
- Acquisti, A., & Grossklags, J. (2007). What can behavioral economics teach us about privacy. In A. Acquisti, S. Gritzalis, C. Lambrinoudakis, & S. di Vimercati (Eds.), *Digital privacy: Theory, technologies and practices, Chapter 18* (pp. 363–377). Auerbach Publications.
- Alaba, F. A., Othman, M., Targio, H., Ibrahim, A., & Alotaibi, F. (2017). Internet of things security: A survey. *Journal of Network and Computer Applications*, 88, 10–28. <https://doi.org/10.1016/J.JNCA.2017.04.002>. ISSN 1084-8045.
- An, B., Brown, M., Vorobeychik, Y., & Tambe, M. (2013). Security games with surveillance cost and optimal timing of attack execution. In *Proceedings of the 12th international conference on autonomous agents and multiagent systems (AAMAS)* (pp. 223–230).
- Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution savings plans. *The American Economic Review*, 91(1), 79–98. <https://www.jstor.org/stable/2677899>.
- Bier, V., Oliveros, S., & Samuelson, L. (2007). Choosing what to protect: Strategic defensive allocation against an unknown attacker. *Journal of Public Economic Theory*, 9(4), 563–587.
- Bleichrodt, H., & Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science*, 46(11), 1485–1496. <https://doi.org/10.1287/mnsc.46.11.1485.12086>. ISSN 0025-1909.
- Bloch, F., Dutta, B., & Dziubinski, M. (2020). A game of hide and seek in networks. [arXiv:abs/2001.03132](https://arxiv.org/abs/2001.03132).
- Boche, H., Naik, S., & Alpcan, T. (2011). Characterization of convex and concave resource allocation problems in interference coupled wireless systems. *IEEE Transactions on Signal Processing*, 59(5), 2382–2394.
- Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4), 1375–1412. <https://doi.org/10.3982/ECTA7139>. ISSN 0012-9682.
- Caplin, A., & Leahy, J. (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1), 55–79. <https://doi.org/10.1162/003355301556347>.
- Caplin, A., & Leahy, J. (2004). The supply of information by a concerned expert. *The Economic Journal*, 114(497), 487–505. <https://doi.org/10.1111/j.0013-0133.2004.0228a.x>.
- Chapman, J., Snowberg, E., Wang, S., & Camerer, C. (2018). Loss attitudes in the U.S. population: Evidence from dynamically optimized sequential experimentation (DOSE). Technical report, National Bureau of Economic Research. <http://www.nber.org/papers/w25072.pdf>.
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97. <https://doi.org/10.1016/J.JBEF.2015.12.001>. ISSN 2214-6350.
- Choi, S., Kim, J., Lee, E., & Lee, J. (2018). *Probability weighting and cognitive ability*. SIER Working Paper Series 121, Institute of Economic Research, Seoul National University.
- Chowdhury, S. M. (2019). The attack and defense mechanisms-Perspectives from behavioral economics and game theory. *Behavioral and Brain Sciences*, 42, e121. <https://doi.org/10.1017/S0140525X19000815>.
- Chowdhury, S. M., Kovenock, D., Rojo Arjona, D., & Wilcox, N. T. (2016). Focality and asymmetry in multi-battle contests. https://digitalcommons.chapman.edu/esi_working_papers/194/.
- Chowdhury, S. M., Kovenock, D., & Sheremeta, R. M. (2013). An experimental investigation of Colonel Blotto games. *Economic Theory*, 52(3), 833–861. <https://doi.org/10.1007/s00199-011-0670-2>. ISSN 09382259.
- Clark, D. J., & Konrad, K. A. (2007). Asymmetric conflict: Weakest link against best shot. *Journal of Conflict Resolution*, 51(3), 457–469. <https://doi.org/10.1177/0022002707300320>.

- Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes*, 38(2), 230–256.
- Deck, C., & Sheremeta, R. M. (2012). Fight or flight?: Defending against sequential attacks in the game of siege. *Journal of Conflict Resolution*, 56(6), 1069–1088. <https://doi.org/10.1177/0022002712438355>.
- Dighe, N. S., Zhuang, J., & Bier, V. M. (2009). Secrecy in defensive allocations as a strategy for achieving more cost-effective attacker deterrence. *International Journal of Performability Engineering*, 5(1), 31–43.
- Djawadi, B. M., Endres, A., Hoyer, B., & Recker, S. (2019). Network formation and disruption—An experiment are equilibrium networks too complex? *Journal of Economic Behavior and Organization*, 157, 708–734. <https://doi.org/10.1016/j.jebo.2018.11.004>. ISSN 01672681.
- Dziubiński, M., & Goyal, S. (2013). Network design and defence. *Games and Economic Behavior*, 79(1), 30–43. <https://doi.org/10.1016/j.geb.2012.12.007>.
- Dziubiński, M., & Goyal, S. (2017). How do you defend a network? *Theoretical Economics*, 12(1), 331–376. <https://doi.org/10.3982/te2088>. ISSN 1555-7561.
- Epper, T., & Fehr-Duda, H. (2018). *Unifying risk taking and time discounting: The missing link*. Economics Working Paper Series 1812, University of St. Gallen, School of Economics and Political Science.
- Fehr-Duda, H., Epper, T., Bruhin, A., & Schubert, R. (2011). Risk and rationality: The effects of mood and decision rules on probability weighting. *Journal of Economic Behavior & Organization*, 78(1–2), 14–24. <https://doi.org/10.1016/J.JEBO.2010.12.004>. ISSN 0167-2681.
- Fehr-Duda, H., de Gennaro, M., & Schubert, R. (2006). Gender, financial risk, and probability weights. *Theory and Decision*, 60(2–3), 283–313. <https://doi.org/10.1007/s11238-005-4590-0>.
- Feng, S., Xiong, Z., Niyato, D., Wang, P., Wang, S. S., & Shen, X. S. (forthcoming). Joint pricing and security investment in cloud security service market with user interdependency. *IEEE Transactions on Services Computing*. <https://www.computer.org/csdl/journal/sc/5555/01/09098048/1k0KZ73ZPmU>.
- Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2014). Game theory meets information security management. In *International information security conference (IFIP)* (pp. 15–29).
- Fréchette, G. R. (2015). Experiments: professionals versus students. In G. Fréchette & A. Schotter (Eds.), *Handbook of experimental economic methodology, Chapter 17* (pp. 360–390). Oxford University Press.
- Frey, B. J. & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315, 972–976. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.3145>.
- Gartner. (2018). Gartner forecasts worldwide information security spending to exceed \$124 Billion in 2019. <https://www.gartner.com/en/newsroom/press-releases/2018-08-15-gartner-forecasts-worldwide-information-security-spending-to-exceed-124-billion-in-2019>.
- Goyal, S., & Vigier, A. (2014). Attack, defence, and contagion in networks. *The Review of Economic Studies*, 81(4), 1518–1542. <https://doi.org/10.1093/restud/rdu013>.
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125. <https://doi.org/10.1007/s40881-015-0004-4>. ISSN 2199-6776.
- Guan, P., He, M., Zhuang, J., & Hora, S. C. (2017). Modeling a multitarget attacker-defender game with budget constraints. *Decision Analysis*, 14(2), 87–107.
- Homer, J., Zhang, S., Ou, X., Schmidt, D., Du, Y., Rajagopalan, S. R., et al. (2013). Aggregating vulnerability metrics in enterprise networks using attack graphs. *Journal of Computer Security*, 21(4), 561–597. <https://doi.org/10.3233/JCS-130475>.
- Hota, A. R., Clements, A. A., Sundaram, S., & Bagchi, S. (2016). *Optimal and game-theoretic deployment of security investments in interdependent assets* (pp. 101–113). Springer. https://doi.org/10.1007/978-3-319-47413-7_6.
- Hota, A. R., Clements, A. A., Bagchi, S., & Sundaram, S. (2018). A game-theoretic framework for securing interdependent assets in networks. In S. Rass & S. Schauer (Eds.), *Game theory for security and risk management: From theory to practice* (pp. 157–184). Springer. https://doi.org/10.1007/978-3-319-75268-6_7.
- Hoyer, B., & Rosenkranz, S. (2018). Determinants of equilibrium selection in network formation: An experiment. *Games*, 9(4), 89. <https://doi.org/10.3390/g9040089>. ISSN 2073-4336.
- Humayed, A., Lin, J., Li, F., & Luo, B. (2017). Cyber-physical systems security—A survey. *IEEE Internet of Things Journal*, 4(6), 1802–1831. <https://doi.org/10.1109/JIOT.2017.2703172>. ISSN 2327-4662.

- Jauhar, S., Chen, B., Temple, W. G., Dong, X., Kalbarczyk, Z., Sanders, W. H., & Nicol, D. M. (2015). Model-based cybersecurity assessment with NESCOR smart grid failure scenarios. In *2015 IEEE 21st Pacific Rim international symposium on dependable computing (PRDC)*. IEEE. <https://doi.org/10.1109/PRDC.2015.37>. ISBN 978-1-4673-9376-8.
- Kosfeld, M. (2004). Economic networks in the laboratory: A survey. *Review of Network Economics*, 3(1), 20–42.
- Kovenock, D., & Roberson, B. (2018). The optimal defense of networks of targets. *Economic Inquiry*, 56(4), 2195–2211. <https://doi.org/10.1111/ecin.12565>.
- Kovenock, D., Roberson, B., & Sheremeta, R. M. (2019). The attack and defense of weakest-link networks. *Public Choice*, 179(3–4), 175–194. <https://doi.org/10.1007/s11127-018-0618-1>. ISSN 15737101.
- Lee, E. (2015). The past, present and future of cyber-physical systems: A focus on models. *Sensors*, 15(3), 4837–4869. <https://doi.org/10.3390/s150304837>. ISSN 1424-8220.
- Leibowitz, H., Piotrowska, A. M., Danezis, G., & Herzberg A. (2019). No right to remain silent: Isolating malicious mixes. In *28th USENIX security symposium (USENIX security 19)* (pp. 1841–1858). USENIX Association. ISBN 978-1-939133-06-9.
- George, L. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, 97(387), 666. <https://doi.org/10.2307/2232929>.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>. ISSN 07495978.
- McBride, M., & Hewitt, D. (2013). The enemy you can't see: An investigation of the disruption of dark networks. *Journal of Economic Behavior & Organization*, 93, 32–50. <https://doi.org/10.1016/j.jebo.2013.07.004>. ISSN 01672681.
- McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1), 6–38. <https://doi.org/10.1006/GAME.1995.1023>.
- Mersinas, K., Hartig, B., Martin, K. M., & Seltzer, A. (2016). Are information security professionals expected value maximizers?: An experiment and survey based test. *Journal of Cybersecurity*, 2(1), 57–70. <https://doi.org/10.1093/cybsec/tyw009>.
- Modelo-Howard, G., Bagchi, S., & Lebanon, G. (2008). Determining placement of intrusion detectors for a distributed application through Bayesian network modeling. In *11th international symposium on research in attacks, intrusions and defenses (RAID)* (pp. 271–290).
- Nguyen, K. C., Alpcan, T., & Basar, T. (2010). Stochastic games for security in networks with interdependent nodes. [arXiv:abs/1003.2440](https://arxiv.org/abs/1003.2440).
- Nikoochal, M. E., & Zhuang, J. (2012). Robust allocation of a defensive budget considering an attacker's private information. *Risk Analysis: An International Journal*, 32(5), 930–943.
- Nithyanand, R., Starov, O., Zair, A., Gill, P., & Schapira, M. (2016). Measuring and mitigating AS-level adversaries against Tor. In *Network & Distributed System Security Symposium (NDSS)*.
- Pal, R., & Golubchik, L. (2010). Analyzing self-defense investments in internet security under cyber-insurance coverage. In *2010 IEEE 30th international conference on distributed computing systems* (pp. 339–347). IEEE.
- Paté-Cornell, M. E., Kuypers, M., Smith, M., & Keller, P. (2018). Cyber risk management for critical infrastructure: A risk analysis model and three case studies. *Risk Analysis*, 38(2), 226–241. <https://doi.org/10.1111/risa.12844>. ISSN 15396924.
- Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3), 497. <https://doi.org/10.2307/2998573>. ISSN 00129682.
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4), 323–343. [https://doi.org/10.1016/0167-2681\(82\)90008-7](https://doi.org/10.1016/0167-2681(82)90008-7). ISSN 0167-2681.
- Sheremeta, R. M. (2019). The attack and defense games. *Behavioral and Brain Sciences*, 42, e140. <https://doi.org/10.1017/S0140525X19000931>. ISSN 0140-525X.
- Sheyner, O., & Wing, J. (2003). Tools for generating and analyzing attack graphs. In *International symposium on formal methods for components and objects (FMCO)* (pp. 344–371). Springer. https://doi.org/10.1007/978-3-540-30101-1_17.
- Sun, X., Shen, C., Chang, T.-H., & Zhong, Z. (2018). Joint resource allocation and trajectory design for UAV-aided wireless physical layer security. In *2018 IEEE Globecom workshops (GC Wkshps)* (pp. 1–6). IEEE.

- Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, 100(1), 557–571. <https://doi.org/10.1257/aer.100.1.557>. ISSN 0002-8282.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323. <https://doi.org/10.1007/BF00122574>. ISSN 0895-5646.
- Wu, D., Xiao, H., & Peng, R. (2018). Object defense with preventive strike and false targets. *Reliability Engineering & System Safety*, 169, 76–80.
- Xie, P., Li, J. H., Xinming, O., Liu, P., & Levy, R. (2010). Using Bayesian networks for cyber security analysis. In *Proceedings of the international conference on dependable systems and networks (DNS)* (pp. 211–220). <https://doi.org/10.1109/DSN.2010.5544924>. ISBN 9781424475018.
- Yang, R., Kiekintveld, C., Ordonez, F., Tambe, M., & John, R. (2011). Improving resource allocation strategy against human adversaries in security games. In *22nd international joint conference on artificial intelligence (IJCAI)*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.