UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Security Economics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Colin Xavier Deurlington

Dissertation Committee:
Professor Stergios Skaperdas, Chair
Professor Michael McBride
Professor Daniel Bogart

2024

DEDICATION

To my parents and siblings:

Thank you for everything you have done for me. Completing this PhD without your constant love and support would have been a task taller than all the mountains we've hiked.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank each of my advisors for their roles in helping me complete this dissertation.

My committee chair, Stergios Skaperdas, introduced me to conflict and security economics, and was instrumental in every step of my doctoral pursuit. His mentorship and friendship were integral to my time here at UCI, and his excitement in current events relating to my research was motivating. I am also grateful to Stergios for the teaching opportunities and many recommendation letters that provided me with funding and conference experiences that were essential to my growth as a researcher and person.

Chapter 2 of this dissertation would not have been possible without the collaboration and mentorship from my co-author, Michael McBride. I greatly appreciate the strong work-ethic and approach to research that was instilled in me during the co-authoring of this paper. He not only provided funding for our experiment, but also spent countless hours with me discussing the direction and results of this paper. Co-authoring with Mike was one of the most valuable learning experiences in my academic career.

I am immensely thankful for Daniel Bogart's thoughtful advice and support, and I especially appreciate his guidance in the development of Chapter 3. Conversations with Dan were crucial when I found myself stuck, and I came out of these discussions inspired and full of new ideas. I look forward to advancing my research in Chapter 3, and I know I can always reach out to Dan for exceptional feedback.

I will forever be grateful to each member of my committee for their guidance and never-ending encouragement throughout this process.

Further, funding from the UCI Economics Department, the UCI School of Social Sciences, and the Center for Global Peace and Conflict Studies (CGPACS) was greatly appreciated and essential to my completion of this degree.

Lastly, I want to acknowledge and thank Eric Fisher, Stephen Hamilton, and Eduardo Zambrano in the integral roles they played in sparking my interest in an economics doctoral degree.

VITA

Colin Xavier Deurlington

EDUCATION

Doctor of Philosophy in Economics **2024**
University of California Irvine *Irvine, California*

Master of Arts in Economics **2022**
University of California Irvine *Irvine, California*

Bachelor of Science in Quantitative Economics **2020**
California Polytechnic State University *San Luis Obispo, California*

RESEARCH FIELDS

Game Theory, Experimental Economics, Political Economics

TEACHING EXPERIENCE

Teaching Assistant **2020–2024**
University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

Essays in Security Economics

By

Colin Xavier Deurlington

Doctor of Philosophy in Economics

University of California, Irvine, 2024

Professor Stergios Skaperdas, Chair

This dissertation provides theoretical, experimental, and empirical studies of topics important in security economics. Chapter 1 and 3 assess cybersecurity settings, in particular, and approaches that limit the frequency of incidents experienced. Chapter 2 provides an experimental justification of Conjectural Equilibrium, an important equilibrium concept particularly relevant to security environments where feedback is limited.

Chapter 1 studies a model of weakest-link network defense. In this model, the defender determines the internal accessibility of a valuable asset and allocates defensive resources prior to an attacker's decision to attack. In equilibrium, one of two resource allocations can arise: (1) both the defender and attacker allocate a strictly positive level of resources, or (2) the defender allocates a sufficient level of resources to deter attacks. As the defender's cost-adjusted valuation of an asset increases relative to the attacker, the defender is more willing to increase the internal accessibility of the asset, irrespective of the marginal benefit from increased accessibility. This model provides theoretical foundations for data breach and other cybersecurity settings.

Chapter 2 provides an experimental test of the Conjectural Equilibrium concept in a threshold public good game with limited feedback. Consistent with our predictions, we find evidence that strategy profiles that are Conjectural Equilibria but not Nash Equilibria are more likely

as feedback decreases, and that subjects are more likely to hold incorrect beliefs as feedback decreases. However, the use of Conjectural Equilibrium as a predictive concept is complicated because risk aversion interacts with the feedback treatment, belief convergence occurs at different rates across treatments, and subjects intentionally choose to not maximize payoffs. Overall, our findings support a measured approach to using the Conjectural Equilibrium concept to obtain predictions in limited-feedback settings. These results are especially useful for understanding security settings, where agents often make decisions based on limited feedback.

Chapter 3 empirically examines the relationship between organizations' cybersecurity measures and their experienced level of incident frequency. Cybersecurity is an increasingly relevant concern for governments, businesses, and individuals. However, despite both rising investment in cybersecurity and frequency of cyber incidents, little research has been done to assess this relationship. Using fixed effects regressions over multiple thresholds of incident frequency, this paper identifies staff cybersecurity training, data storage rules, and restrictions on personal devices used for work as measures associated with reduced incident frequency. Furthermore, this paper provides a foundational assessment of how cybersecurity measures are associated differently with phishing versus non-phishing incidents, providing a first step in understanding the usefulness of measures in preventing incidents of different severities.

Chapter 1

Defense and Connectivity of Weakest-Link Networks

1.1 Introduction

In networks, there is a trade-off between efficiency and security (Morselli et al., 2007). Organizations, especially in digital environments, must often determine whether to provide employees, departments, or outside consultants with access to valuable assets, such as databases containing personal, financial, or medical information. Increased access to data can help an organization provide better services and operate more efficiently. As more access is given, however, these databases become increasingly vulnerable to outside threats. Maintaining security is critically important, as data breaches can be quite costly, both in terms of money and reputation.[1]

This paper explores a network defender's decisions in a contest setting resembling a data

[1]See Toulas (2022) for a report on the California Heritage Provider Network Data Breach, which impacted over three million patients in December 2022.

security environment. Specifically, should a network defender increase access to a valuable node when increasing access makes this node more vulnerable? As an illustrative example, consider a database containing financial or medical information, which is valuable both to a firm as well as a potential hacker. The defender must decide whether to give employees within the firm access to this data, which provides the firm with better knowledge of its customers' information and needs. However, providing employees with access to this data makes the data more vulnerable to phishing attacks – more employees with access means more targets for the attacker to phish in an attempt to breach the data. This example can also be applied in an infrastructure setting, where a government agency or municipality determines access to water filtration controls, electrical grid flows, or even nuclear codes. We can think of this setting generally as a simple star network, where the central node is some valuable asset, undirected edges signify access to this asset, and peripheral nodes can be individuals, departments, or consultants that offer value from being able to access the valuable asset.[2]

I use a sequential weakest-link lottery contest framework to address the question of when additional access to a valuable asset should be allotted in an organization.[3] In this framework, a network defender first decides whether to isolate or allow access to their valuable asset. Additional benefit is gained for each node that is given access to the valuable asset, which comes at the cost of increased network vulnerability. The defender then allocates defensive resources to protect her network. I independently consider two defensive settings in this environment – infrastructure defense and individual node defense. In the former defensive setting, the defender contributes resources using a technology that uniformly protects all nodes in her network.[4] In the latter defensive setting, the defender contributes resources to

[2]As my model will be considering a simple star network structure and weakest-link objective function, the use of directed rather than undirected edges makes no qualitative difference.

[3]For more on lottery contests, see Tullock (1980) and Skaperdas (1996).

[4]To my knowledge, this is the first paper that discusses infrastructure defense in a lottery contest setting. Kovenock and Roberson (2012) and Lizzeri and Persico (2001) examine infrastructure defense in auction contest settings.

defend each node individually. After observing the defender's network connection decision and security allocation, an attacker determines where to distribute attack resources across the network. The probability a given node is defended is determined by the defensive resources relative to the total resources allocated to the node. The attacker breaches the network if she successfully attacks at least one node.

This model extends the weakest-link multiple-battlefield lottery contest framework established in Clark and Konrad (2007) in four important ways. First, my model allows for the defender to decide on the level of connectivity of the weakest-link network prior to determining how to allocate defensive resources. Second, I modify the value of victory to the defender to be increasing in the size of the network being defended. These first two extensions are especially important for capturing the incentives of a network defender in a weakest-link setting.[5] The third extension involves using a sequential-move contest framework, where the adversary observes the defender's network size and security infrastructure before deciding where and how much to concentrate attack resources. This assumption has not been applied to a weakest-link network defense setting (Arce et al., 2012; Kovenock and Roberson, 2018; Levitin and Hausken, 2010)[6], yet is common in other network defense literature (Goyal and Vigier, 2014; Dziubiński and Goyal, 2013; Powell, 2009). Finally, I analyze this model under two different defensive setting – infrastructure and individual node defense.

Equilibrium analysis of my model shows two types of resource allocations can arise: (1) both the network defender and attacker allocate a strictly positive level of resources to every node, or (2) the network defender spends a sufficient level of resources so as to deter all attacks on the network. In both equilibrium allocations, the defender allocates a strictly positive level of force to protect her network, offering a different result from that found in Clark and

[5] A defender should only be willing to take on the increased vulnerability from defending larger networks if larger networks provide additional value when successfully defended. Furthermore, if the benefit of more connections does not outweigh the added risk, a defender should limit the size of the space they must defend.

[6] Levitin and Hausken (2010) analyze a repeated contest, where the attacker needs to succeed once to win, though both the defender and attacker allocate resources simultaneously in each iteration of the contest.

Konrad (2007), where the defender surrenders and does not defend her network with some positive probability. To my knowledge, the second equilibrium allocation of deterrence is a novel finding in weakest-link lottery contest research, and provides a valuable theoretical foundation for policies attempting to combat the rapid rise in cyber attacks over the past few years.[7]

Equilibrium analysis also provides two insights into a defender's decision for increasing the internal accessibility of an asset. First, the defender is more likely to increase access to the valuable node when her cost-adjusted asset value is greater than that of the attacker. Though perhaps unsurprising, this result is consistent across both defensive settings, all network sizes, and regardless of the marginal benefit of increasing access. Assuming that the values to the defender and attacker from winning this contest are fairly stable over time, this indicates that the decision to increase the accessibility in a network is highly dependent on the relative costs of defending to attacking resources. Second, when the defender is deciding whether to increase access to a valuable asset, the individual benefit from these additional connections generally does not need to be as great when more connections are being added. That is to say, the value added from granting ten moderately adept analysts access to confidential data may be worth the additional risk of the data being stolen while the value added from granting access to only five more skilled analysts may not.

This paper contributes primarily to research in contests involving a defender with a weakest-link objective and an attacker with a best-shot objective. Clark and Konrad (2007) introduces a simple multiple-battlefield lottery contest where the defender must successfully defend each battlefield. Levitin and Hausken (2010) extends this model to allow the attacker to decide whether to fight over the battlefields one at a time or all in one period.[8] In multiple-

[7]Though not precisely the setting studied in this paper, ransomware and data security are similar problems. See Financial Crimes Enforcement Network (2021) for a report on how the number and cost of ransomware attacks tripled from 2020 to 2021.

[8]Klumpp et al. (2019) analyzes a similar repeated multiple-battlefield lottery contest to understand electoral competitions.

battlefield lottery contests, the attacker uniformly allocates resources across the battlefields (Clark and Konrad, 2007; Kovenock and Roberson, 2012), whereas the equilibrium attack strategy in auction contest settings involves only targeting one battlefield (Arce et al., 2012; Kovenock and Roberson, 2018). Arce et al. (2012) considers the possibility of multiple attack technologies to model terrorism and understand when an attacker decides to strategically use suicide attacks. Weakest-link contests have also been studied in various experimental settings (Kovenock et al., 2019; Deck and Sheremeta, 2012).[9] For a survey of results in other simultaneous-move multiple-battlefield conflicts, with weakest-link or other asymmetric objective functions, see Kovenock and Roberson (2012).[10]

Directly related to multiple-battlefield contests, much has been done in studying two-player simultaneous resource allocation settings in the Colonel Blotto game, originating with Borel (1921). Roberson and Kvasov (2012) consider the players' incentives in an all-pay auction framework with asymmetric budget constraints. Kovenock and Roberson (2021) characterize Nash equilibrium for both Colonel Blotto and General Lotto games where the battlefield values are heterogeneous and asymmetric across players, finding that players benefit from targeting battlefields where their relative valuations are lower. Chowdhury et al. (2013) and Montero et al. (2016) offer experimental investigations of the theoretical results in Blotto contests, finding evidence that players tend to allocate a higher concentration of resources to fewer battlefields than equilibrium analysis predicts.

My model also contributes to research in sequential-move network defense – specifically, where the attacker allocates resources after the defender establishes their network structure and defensive allocations. Powell (2009) provides a concise equilibrium analysis of a general lottery contest in this sequential-move setting. Goyal and Vigier (2014) aims to find the

[9]Kovenock et al. (2019) finds that lab participants do not follow the equilibrium behavior expected of attackers in lottery contests. A likely reason for this is the tendency for individuals to use focal points rather than optimal strategies when making their allocation of attack resources (Chowdhury et al., 2016).

[10]For a recent simultaneous move model that incorporates asymmetries and heterogeneous values across battlefields, see also Li and Zheng (2022).

ideal network structure and defensive allocation while Acemoglu et al. (2016) explores the implications of decentralized defense in models of contagion. Dziubiński and Goyal (2013, 2017) add to the discussion on optimal network design in the face of an adversary, where the former considers node-removal attacks rather than attack resource allocations and the latter considers decentralized defense.[11] Hausken (2006, 2017) and Hota et al. (2018) study sequential-move models where successfully attacking one node in a network influences the probability of successful attacking connected nodes.[12] Hausken (2017) considers a simple case of only two interdependent nodes, while Hausken (2006) and Hota et al. (2018) consider settings of decentralized defense.

The remainder of this article proceeds as follows. Section 2 describes the model. Section 3 solves for the equilibrium allocations and profits in both infrastructure and individual node defense settings. The defender's decision to connect their network is also solved for in this section under both defensive settings. Section 4 discusses and concludes.

1.2 Model

I consider a sequential weakest-link network defense model. Specifically, the defender initially decides on the size of the network and the defensive allocations, and after observing these choices, the attacker decides her attack allocations. My model is formally established below.

2.1 – Players/setting

There are two agents in this contest – a network defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$. Define $N = \{1, \ldots, n\}$ as the set of nodes $\mathcal{D}$ defends.[13] Suppose only one of these nodes is

[11] For more on research investigating node-removal attacks, see Bloch et al. (2020) and McBride and Hewitt (2013). These papers analyze settings of incomplete information, where a government organization must remove a node to disrupt or capture a key member in a criminal network.

[12] The extreme case where a successful attack at one node implies successful attacks at all adjacent nodes is analogous to a weakest-link network.

[13] These can also be conceptualized as "battlefields" or "targets," but I will be using "nodes" throughout

considered directly valuable, or "high-value," while the remaining $n-1$ nodes are only indirectly valuable, or "low-value." As an illustrative example, suppose the defender is a business owner and the attacker is some hacker. The business collects consumer financial information and compiles it in a large database – this database is a "high-value" node, as both the business owner and the hacker value the information contained within the database. The data becomes more valuable to the business owner, though, when she allows employees or consultants to assess and work with the data – the individuals with access to the database are "low-value" nodes, as they only provide value to the business owner when able to access the data. However, as the number of connections to the database increases, the information becomes easier for a hacker to gain access to and steal.

2.2 – Moves

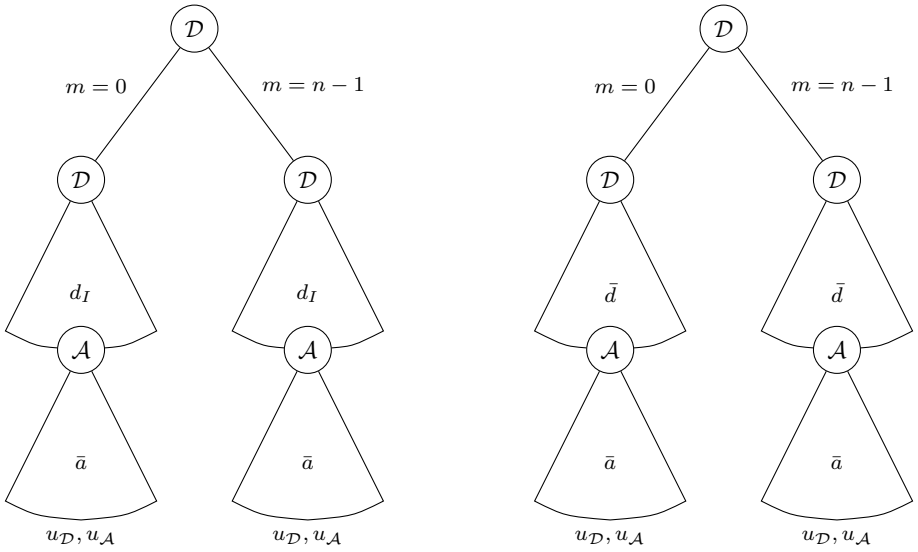
Agents move sequentially. As the first mover, $\mathcal{D}$ makes two decisions. First, $\mathcal{D}$ decides whether to connect low-value nodes to her high-value node. In this paper, I consider a simple binary choice: $\mathcal{D}$ either isolates her high-value node or connects all her low-value nodes to the high-value node. This is to provide a baseline understanding of the edge cases. Define m as the choice variable for how many low-value nodes $\mathcal{D}$ connects to the high-value node, so that $m \in \{0, n-1\}$. More explicitly, $\mathcal{D}$ decides between defending a network of size 1 or of size n. Define C as the component containing the high-value node, with $|C| = m+1$.

$\mathcal{D}$'s second decision involves how to allocate defensive resources across her network. I consider two settings: (1) defense under an infrastructure technology, $d_I > 0$, and (2) individual node defense, $\bar{d} = (d_1, \ldots, d_n)$. In the former setting, $\mathcal{D}$ allocates resources to an infrastructure technology, d_I, that uniformly defends her entire network. In the latter setting, $\mathcal{D}$ allocates resources to defend each node individually. To my knowledge, the presence of an infrastructure technology has not been studied in a lottery contest setting. This is a reasonable notion for a cybersecurity setting, however, as firewalls, training, or

this paper to refer to the possible points where $\mathcal{D}$ and $\mathcal{A}$ may allocate resources.

other monitoring/scanning tools arguably protect the entire network for one overall cost to the network defender. Individual node defense is a standard approach in network security and multiple-battlefield contests, and therefore provides a baseline framework to compare with the results offered from including an infrastructure technology.

As the second mover, $\mathcal{A}$ observes $\mathcal{D}$'s actions and allocates attack resources, $\bar{a} = (a_1, \ldots, a_n)$, across the network. This sequence of moves is shown concisely in Figure 1.1:



(a) infrastructure defense, d_I

(b) individual node defense, $\bar{d} = (d_1, \ldots, d_n)$

Figure 1.1: The sequential game

2.3 – Contest Success Function

The winner at each node is determined by a standard Tullock contest success function (Tullock, 1980). Define the probability that $\mathcal{D}$ successfully defends node j as:

$$p_\mathcal{D}(d_j, a_j) = \begin{cases} \dfrac{d_j}{d_j + a_j} & if\ a_j > 0 \\ 1 & otherwise \end{cases}^{14} \tag{1.1}$$

Notice that $d_j = d_I$ for all j under infrastructure defense.

The winner of the overall contest is determined by weakest-link and best-shot objective functions over the component containing the high-value node for $\mathcal{D}$ and $\mathcal{A}$, respectively. That is, $\mathcal{D}$ wins the contest if she successfully defends every node within the component containing the high-value node, and $\mathcal{A}$ wins the contest if she successfully attacks at least one node within the component containing the high-value node.

2.4 – Node values

A high-value node provides a direct value of $v_{\mathcal{D}} > 0$ to $\mathcal{D}$ when successfully defended and a direct value of $v_{\mathcal{A}} > 0$ to $\mathcal{A}$ when successfully attacked. A low-value node that is not connected to the high-value node provides no benefit to either $\mathcal{D}$ or $\mathcal{A}$. However, connecting a low-value node to a high-value node provides a value of $\beta v_{\mathcal{D}}$, $\beta > 0$, to $\mathcal{D}$. By the objective functions (weakest-link and best-shot), the network is more vulnerable to attack when low-value nodes are connected to the high-value node. To be precise, $\mathcal{D}$ receives a value of $V_{\mathcal{D}} = v_{\mathcal{D}}(1 + m\beta)$ for connecting m low-value nodes to the high-value node and successfully defending the connected network. $\mathcal{A}$, on the other hand, receives a value of $V_{\mathcal{A}} = v_{\mathcal{A}}$ for successfully attacking a single node in the component containing the high-value node.

2.5 – Costs

I assume linear cost functions for both agents. In the case of infrastructure defense, $\mathcal{D}$'s cost function is simply $c(d_I) = d_I$. In the case of individual node defense, $\mathcal{D}$'s cost function is $c(\bar{d}) = \sum_{j=1}^{n} d_j$. Under both settings, $\mathcal{A}$'s cost function is $c(\bar{a}) = \sum_{j=1}^{n} \alpha a_i$, where $\alpha > 0$ is the cost of attack resources relative to defensive resources. Intuitively, this α captures how costly it is for $\mathcal{D}$ to implement multi-factor authentication and for $\mathcal{A}$ to purchase technical exploits on the dark web, for example.

[14] This assumption follows other network defense frameworks (Clark and Konrad, 2007; Goyal and Vigier, 2014).

2.6 – Payoff functions

Compiling the above information, each agent's expected payoff from the contest is simply the product of their probability of winning the contest and their value from winning minus the cost of their resource allocation. I present $\mathcal{D}$ and $\mathcal{A}$'s expected payoff functions in the case of a defensive infrastructure technology, d_I, in equations 1.2 and 1.3.

$\mathcal{D}$'s expected payoff function under a defensive infrastructure technology:

$$\begin{aligned} u_{\mathcal{D}} &= \prod_{i=1}^{m+1} p_{\mathcal{D}}(d_I, \bar{a}) V_{\mathcal{D}} - c(d_I) \ , \quad m \in \{0, n-1\} \\ &= \left(\prod_{i=1}^{m+1} \left(\frac{d_I}{d_I + a_j(d_I)} \right) \right) v_{\mathcal{D}}(1 + m\beta) - d_I, \quad m \in \{0, n-1\} \end{aligned} \tag{1.2}$$

$\mathcal{A}$'s expected payoff function under a defensive infrastructure technology:

$$\begin{aligned} u_{\mathcal{A}} &= \left(1 - \prod_{i=1}^{m+1} p_{\mathcal{D}}(d_I, \bar{a}) \right) V_{\mathcal{A}} - c(\bar{a}) \ , \quad m \in \{0, n-1\} \\ &= \left(1 - \prod_{i=1}^{m+1} \left(\frac{d_I}{d_I + a_j} \right) \right) v_{\mathcal{A}} - \alpha \sum_{i=1}^{m+1} a_i, \quad m \in \{0, n-1\} \end{aligned} \tag{1.3}$$

The agents' expected payoff functions in the case of individual node defense, $\bar{d} = (d_1, \ldots, d_n)$, is provided at the beginning of Section 3.2 in equations 1.4 and 1.5.

2.7 – Summary

My paper modifies the two-agent weakest-link multiple-battlefield contest framework in Clark and Konrad (2007) in four ways. First, I allow the defender to decide between defending a single node or a network of n nodes. Second, I modify the value to the defender of a successful defense to be increasing in the number of battlefields. Though the value to the defender of a successfully defense is increasing in n, the vulnerability of the weakest-link network is also increasing in n. This captures the idea of the efficiency versus security

trade-off in observed in networks (Morselli et al., 2007; Krebs, 2002). Third, I consider a sequential move framework similar to Powell (2009) and Goyal and Vigier (2014), where the attacker allocates resources to battlefields only after observing the defender's network structure and defensive allocations. Finally, I solve this model for both a setting with a defensive infrastructure technology and a setting where each node is defended individually.

In the following section, I solve for the equilibrium resource allocations in both the infrastructure and individual node defense settings. I then analyze $\mathcal{D}$'s equilibrium decision in each setting to connect their network or isolate the high-value node (i.e. $m = n - 1$ or $m = 0$).

1.3 Equilibrium Results and Analysis

In this section, I first solve for the equilibrium allocations of the model described above and analyze $\mathcal{D}$'s decision to connect or isolate the high-value node. I then consider the same model with individual node defense, $\bar{d} = (d_1, \ldots, d_n)$, rather than infrastructure defense to better understand the impact of an infrastructure technology in the lottery contest framework. The following result in Lemma 1 arises in many network defense settings, though it is essential to establish immediately to avoid confusion while discussing later results.

Lemma 1. *If node i is not in the component containing the high-value node ($i \notin C$), then $\mathcal{A}$ does not allocate resources to attack i ($a_i^* = 0$).*

Proof. Proof in Appendix A. □

Due to the result in Lemma 1 and for notational and argumentative simplicity, the remainder of my analysis will only consider the nodes in the component containing the high-value node. That is, all of the equilibrium allocations and analysis in this section pertain only to nodes

$i \in C$. Furthermore, in Section 3.2, I consider the setting where $\mathcal{D}$ allocates resources to individual nodes ($\bar{d} = (d_1, \ldots, d_n)$). A similar trivial proof can be shown that $d_i^* = 0$ if $i \notin C$.

1.3.1 Infrastructure Defense

Equilibrium Allocations

Before solving for $\mathcal{D}$ and $\mathcal{A}$'s equilibrium allocations, it is essential to discuss two initial lemmas. The first of these establishes that $\mathcal{A}$ distributes resources symmetrically across the weakest-link network containing the high-value node:

Lemma 2. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. If $\mathcal{D}$ allocates defensive resources with an infrastructure technology, d_I, and $\mathcal{A}$ allocates attack resources to each node individually, $\bar{a} = (a_1, \ldots, a_n)$), $\mathcal{A}$ will optimally distribute resources uniformly across all nodes in the component containing the high-value node ($a_i^* = a_j^* = a^* \ \forall \ i, j \in C$).*

Proof. Proof in Appendix A. □

A successful attack only requires that $\mathcal{A}$ is victorious at one of the nodes in C. Furthermore, because there do not exist asymmetries across nodes, the marginal benefit from attacking each node is equivalent.

The second important initial lemma establishes that $\mathcal{A}$ is deterred from attacking if $\mathcal{D}$ has a sufficient level of resources defending the network:

Lemma 3. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. Let*

$\mathcal{D}$ allocate defensive resources with an infrastructure technology, d_I, and $\mathcal{A}$ allocate attack resources to each node individually, $\bar{a} = (a_1, \ldots, a_n)$). $\mathcal{A}$ does not attack if $\mathcal{D}$ allocates sufficiently high defensive infrastructure $\left(a^* = 0 \; if \; d_I^* \geq \frac{v_{\mathcal{A}}}{\alpha}\right)$. $\mathcal{D}$ does not allocate defensive resources beyond the point at which $\mathcal{A}$ does not attack $\left(d_I^* \leq \frac{v_{\mathcal{A}}}{\alpha} \; if \; a^* = 0\right)$.

Proof. Proof in Appendix A. □

This occurs when the risk of an unsuccessful attack, and therefore the waste of effort or resources, outweighs the potential prize to $\mathcal{A}$ from capturing the valuable asset. For example, a cybercriminal should not spend time or use sophisticated tools attempting to breach the network of a security-conscious business that does not offer much value to the attacker. $\mathcal{A}$ would be better off targeting firms with worse security or more valuable assets (banks, hospitals, government agencies/municipalities, etc.).

Let $\nu = \frac{v_{\mathcal{D}}}{v_{\mathcal{A}}/\alpha}$ represent the ratio of cost-adjusted values of the high-value node for $\mathcal{D}$ and $\mathcal{A}$. This will be used throughout the remainder of this paper for notational convenience.

In the following proposition, I introduce the equilibrium allocations for $\mathcal{D}$ and $\mathcal{A}$. The first equilibrium allocation corresponds to a state where $\mathcal{D}$ provides full security and as a result, $\mathcal{A}$ does not bother attacking the network. The second equilibrium allocation corresponds to a state where $\mathcal{D}$ does not allocate enough resources to deter $\mathcal{A}$ from attacking the network.

Proposition 1. *Consider a model of sequential network defense between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$. $\mathcal{D}$ connects $m \in \{0, n-1\}$ low-value nodes to a high-value node and allocates infrastructure defense, d_I, that uniformly protects the network. $\mathcal{A}$ observes $\mathcal{D}$'s choices and allocates attack resources $\bar{a} = (a_1, \ldots, a_n)$ to each node. $\mathcal{D}$ has a weakest-link objective, receives $v_{\mathcal{D}}(1 + m\beta)$ from a successful defense, and has linear costs. $\mathcal{A}$ has a best-shot objective, receives $v_{\mathcal{A}}$ from a successful attack, and has linear costs with the relative price of attack resources to defense resources represented by α. The outcome at each node is*

determined by a lottery contest success function. Given the exogeneous parameters, there exist two types of subgame perfect equilibrium allocations that can arise: Attack *or* Deterrence.

1. **Attack:** $\mathcal{A}$ *uniformly allocates a positive level of attack resources across the component containing the high-value node* ($a_i^* = a^* > 0 \ \forall \ i \in C$). *If* $\nu < \dfrac{m+2}{(m+1)(1+m\beta)}$, $\mathcal{D}$ *and* $\mathcal{A}$*'s subgame perfect equilibrium allocations,* $(d_I^* \ , \ a^* \ \forall \ i)$, *are:*

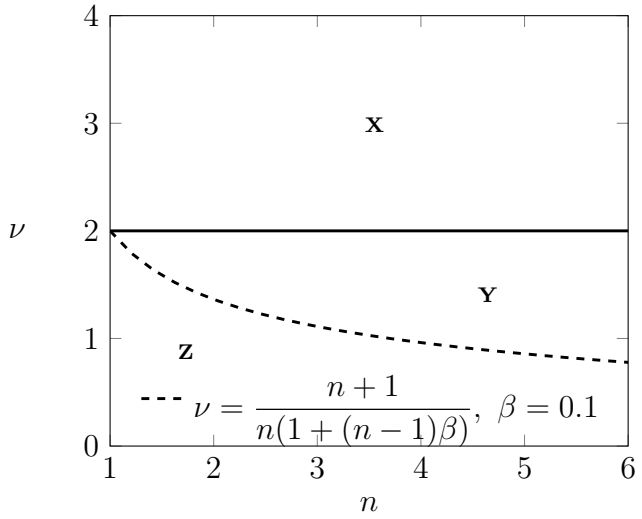
$$d_I^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2}, \ \textit{and}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1 \right].$$

2. **Deterrence:** $\mathcal{D}$ *allocates a sufficiently high level of defensive infrastructure such that* $\mathcal{A}$ *does not allocate resources to attacking* $\mathcal{D}$*'s network* ($a_i^* = a^* = 0 \ \forall \ i$). *If* $\nu \geq \dfrac{m+2}{(m+1)(1+m\beta)}$, $\mathcal{D}$ *and* $\mathcal{A}$*'s subgame perfect equilibrium allocations,* $(d_I^* \ , \ a^* \ \forall \ i)$, *are:*

$$d_I^* = \frac{v_{\mathcal{A}}}{\alpha}, \ \textit{and}$$
$$a^* = 0.$$

Proof. Proof in Appendix A. □

Figure 1.2 below shows the regions corresponding to each type of equilibrium. Note that I use $\beta = 0.1$ to help illustrate, however for all values of $\beta > 0$, each region exists and $\nu = \dfrac{n+1}{n(1+(n-1)\beta)}$ is decreasing and convex. As β increases (decreases), the area of region Z decreases (increases). This follows our intuition – as β increases, $\mathcal{D}$ receives a higher payoff from connecting nodes to her high-value node and therefore has more incentive to deter $\mathcal{A}$ from attacking her network when it is fully connected.



X – Deterrence for $m \in \{0, n-1\}$
Y – Deterrence for $m = n-1$, Attack for $m = 0$
Z – Attack for $m \in \{0, n-1\}$

Figure 1.2: Infrastructure Defense: Equilibrium Regions

The first point to note about Proposition 1 is that if $\mathcal{D}$ has a high cost-adjusted value relative to $\mathcal{A}$, she is willing to fully fund security infrastructure to disincentivize $\mathcal{A}$ from attacking. In region X, $\mathcal{D}$ will allocate a sufficient level of defensive infrastructure to deter $\mathcal{A}$ from attacking regardless of $\mathcal{D}$'s decision to increase the size of her network. While $\nu = 2$ seems like a surprisingly low threshold at which point $\mathcal{D}$ deters all attacks, it is possible that the relative price of attacking resources to defensive resources, α, is remarkably small, making $\nu \geq 2$ possibly unlikely to be observed in the real world.[15]

A second important point is that for a moderate value of ν (region Y), $\mathcal{D}$ fully funds security infrastructure only when her network is connected. If she chooses to isolate the high-value node, however, then she will not allocate sufficient defensive resources to deter $\mathcal{A}$ from attacking. Notice that the area of region Y increases as β increases. This means region Y can be thought of as a setting where $\mathcal{D}$ values the connectivity of her network (or alternatively,

[15]For example, sending phishing emails is a relatively inexpensive method of attacking an organization's network.

the payoff gained from providing access to the high-value node) relatively more than the high-value node itself.

Finally, region Z indicates a setting where $\mathcal{D}$ has a low cost-adjusted value of the high-value node relative to $\mathcal{A}$. When this is the case, it follows that $\mathcal{D}$ will not have a strong desire to invest in security to the point of deterring all attacks from $\mathcal{A}$. Furthermore, $\mathcal{A}$ is willing to spend more resources attacking the network due to the relatively high prize she receives from gaining access to the high-value node. The area of region Z is larger when there are less gains from increasing access to the high-value node (small β).

In order to further explore these equilibria, it is important to find the associated equilibrium payoffs. I provide the payoffs to $\mathcal{D}$ and $\mathcal{A}$ below with respect to being in the "Deterrence" equilibrium ($d_I^* = v_{\mathcal{A}}/\alpha$) or the "Attack" equilibrium ($d_I^* \neq v_{\mathcal{A}}/\alpha$).

Proposition 2. *Based on the subgame perfect equilibrium allocations in Proposition 1, where $\mathcal{D}$ allocates defensive resources with an infrastructure technology, d_I, the subgame equilibrium profits for $\mathcal{D}$ and $\mathcal{A}$, respectively, are*

$$u_{\mathcal{D}}^* = \begin{cases} (1+m\beta)v_{\mathcal{D}} - \dfrac{v_{\mathcal{A}}}{\alpha} & if\ d_I^* = \dfrac{v_{\mathcal{A}}}{\alpha} \\ \left(\dfrac{1+m\beta}{m+2}\right)^{m+2} (\nu(m+1))^{m+1} v_{\mathcal{D}} & if\ d_I^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

$$u_{\mathcal{A}}^* = \begin{cases} 0 & if\ d_I^* = \dfrac{v_{\mathcal{A}}}{\alpha} \\ v_{\mathcal{A}} \left\{ 1 - \nu^m \left(\dfrac{(m+1)(1+m\beta)}{m+2}\right)^{m+1} \left[\nu - (m+1) + (m+1)\nu^2 \left(\dfrac{(m+1)(1+m\beta)}{m+2}\right) \right] \right\} & if\ d_I^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right), \end{cases}$$

where $d_I^* = \dfrac{v_{\mathcal{A}}}{\alpha}$ in the "Deterrence" subgame equilibrium and $d_I^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right)$ in the "Attack" subgame equilibrium.

Proof. Proof in Appendix A. □

A noticeable feature of these equilibrium payoffs is that $u_{\mathcal{D}}$ is increasing in the size of the network, n, when in the "Deterrence" equilibrium. This result comes from the assumptions

in this model that infrastructure defense allows $\mathcal{D}$ to defend a single node at the same cost as n nodes and that there is no cost of maintaining connections. Additionally, it should be pointed out that under both the "Deterrence" and "Attack" equilibria, $\mathcal{D}$ receives a strictly positive payoff. In the remainder of this section, I use these equilibrium payoffs to analyze $\mathcal{D}$'s connection decision. That is, when should $\mathcal{D}$ increase the size of her network, and when should she isolate her valuable assets?

Connection Decision

$\mathcal{D}$ initially decides to connect $m = n - 1$ nodes to a high-value node or to isolate the high-value node ($m = 0$). That is, $\mathcal{D}$ chooses to defend a network of size n or of size 1. $\mathcal{D}$ should connect if $u^*_{\mathcal{D}}|_{m=n-1} \geq u^*_{\mathcal{D}}|_{m=0}$. Below, I explore $\mathcal{D}$'s choice for each of the equilibrium regions shown in Figure 1.2.

Region X:

By Proposition 1, region X corresponds to $\mathcal{A}$ being deterred from attacking both connected and isolated networks. That is,

$$(d^*_I,\ a^*\ \forall\ i) = \left(\frac{v_{\mathcal{A}}}{\alpha},\ 0\right) \text{ for } m \in \{0, n-1\}$$

Therefore, because $u_{\mathcal{D}}$ is strictly increasing in n when $a^* = 0$, $\mathcal{D}$ connects her network. This can also be shown below:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{X}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{X}}|_{m=0} \\ (1 + (n-1)\beta)v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} &\geq v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} \\ (n-1)\beta &\geq 0 \end{aligned}$$

This always holds, so $\mathcal{D}$ always connects in region X. This is a fairly unsurprising result given

the model framework. Because there are no additional costs for $\mathcal{D}$ to defend and maintain a larger network, and because $u_{\mathcal{D}}$ is increasing in β, $\mathcal{D}$ will want as large of a network as possible.

Region Y:

Region Y is more tricky. By Proposition 1, if $\mathcal{D}$ decides to fully connect her network, she will spend sufficient defensive resources to deter $\mathcal{A}$ from attacking. However, if $\mathcal{D}$ chooses to isolate her valuable asset, she will not spend sufficient resources to deter $\mathcal{A}$. That is, the equilibrium allocations are:

$$(d_I^*,\ a^*\ \forall\ i) = \begin{cases} \left(\dfrac{\nu v_{\mathcal{D}}}{4},\ \dfrac{(2-\nu)v_{\mathcal{D}}}{4}\right) & \text{for } m = 0 \\ \left(\dfrac{v_{\mathcal{A}}}{\alpha},\ 0\right) & \text{for } m = n-1 \end{cases}$$

Note, $\nu < 2$ in region Y of Figure 1.2, so $a^* > 0$ when $\mathcal{D}$ chooses $m = 0$. Using these equilibrium allocations, I find that $\mathcal{D}$ connects her network when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{Y}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{Y}}|_{m=0} \\ (1+(n-1)\beta)v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} &\geq \frac{\nu v_{\mathcal{D}}}{4} \\ (1+(n-1)\beta)\nu &\geq \frac{\nu^2}{4} + 1 \\ (n-1)\beta &\geq \frac{\nu}{4} + \frac{1}{\nu} - 1 \\ \beta &\geq \frac{1}{n-1}\left(\frac{\nu^2+4-4\nu}{4\nu}\right) \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{n-1}\left(\frac{(\nu-2)^2}{4\nu}\right)$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. That is, if β is sufficiently large, $\mathcal{D}$ prefers a connected network that is fully defended to an isolated network without full defense. Figure 1.3 shows the connection threshold, $\bar{\beta}$, for region Y.

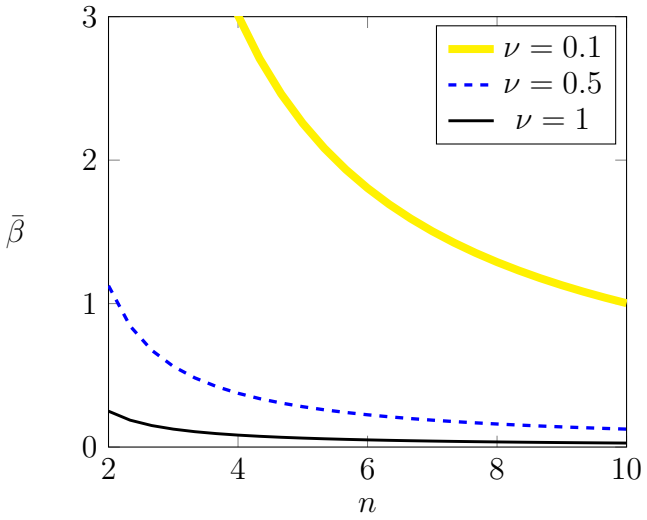


Figure 1.3: Infrastructure Defense: $\mathcal{D}$'s Connection Threshold in Region Y

Notice that $\bar{\beta}$ is decreasing in n and ν (for $\nu \in (0, 2)$). This corresponds to the comparative statics:

$$\frac{\partial \bar{\beta}}{\partial n} = -\frac{1}{(n-1)^2}\left(\frac{(\nu-2)^2}{4\nu}\right) < 0$$
$$\frac{\partial \bar{\beta}}{\partial \nu} = \frac{1}{n-1}\left(\frac{\nu^2-4}{4\nu^2}\right) < 0$$

Again, if this $\bar{\beta}$ threshold is met, $\mathcal{D}$ prefers as large of a network as possible due to the increasing returns of defense and no maintenance cost of connections. Furthermore, if $\mathcal{D}$ has a cost-adjusted value of the high-value node that is large relative to $\mathcal{A}$, she should be willing to fund security to the point of deterring $\mathcal{A}$'s attacks. This implies $\mathcal{D}$ is more likely to connect for greater ν because $\bar{\beta} \approx 0$. However, for lower values of ν, $\mathcal{A}$ has a much higher relative cost-adjusted valuation of gaining access to the component containing the high-value node. As a result, the threshold of security infrastructure required to deter all attacks is substantial, and $\mathcal{D}$ will not be willing to fund this infrastructure unless n or β are sufficiently large.

Region Z:

By Proposition 1, $\mathcal{A}$ allocates a strictly positive level of resources to attack either a connected or isolated network in region Z. The equilibrium allocations for $m \in \{0, n-1\}$ are:

$$\begin{aligned} d_I^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \\ a_i^* = a^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1 \right] \quad \forall\ i \end{aligned}$$

Using these equilibrium allocations, I find that $\mathcal{D}$ connects her network in region Z when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{Z}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{Z}}|_{m=0} \\ \nu^n v_{\mathcal{D}} n^n \left(\frac{1+(n-1)\beta}{n+1} \right)^{n+1} &\geq \frac{\nu v_{\mathcal{D}}}{4} \\ 1+(n-1)\beta &\geq \left(\frac{1}{4\nu^{n-1} n^n} \right)^{1/(n+1)} (n+1) \\ \beta &\geq \frac{1}{n-1} \left[\left(\frac{1}{4\nu^{n-1} n^n} \right)^{1/(n+1)} (n+1) - 1 \right] \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{n-1} \left(\frac{n+1}{(4n^n \nu^{n-1})^{1/(n+1)}} - 1 \right)$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. That is, if β is sufficiently large, $\mathcal{D}$ prefers an imperfectly defended connected network to an imperfectly defended isolated network. Figure 1.4 shows the connection threshold, $\bar{\beta}$, for region Z.

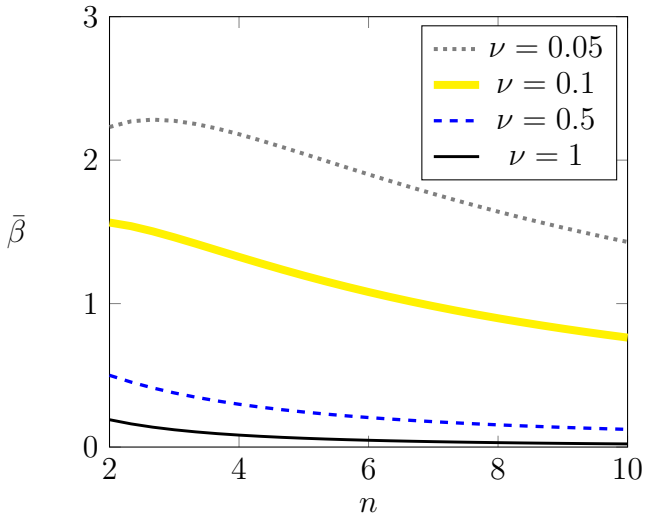


Figure 1.4: Infrastructure Defense: $\mathcal{D}$'s Connection Threshold in Region Z

Similar to region Y, $\bar{\beta}$ is decreasing in ν for $\nu \in (0, 2)$, which corresponds with the following:

$$\begin{aligned}\frac{\partial \bar{\beta}}{\partial \nu} &= \frac{n+1}{(4n^n)^{1/(n+1)}(n-1)} \left(-\frac{n-1}{n+1}\right) \nu^{(n-1)/(n+1)-1} \\ &= -\frac{\nu^{-2/(n+1)}}{(4n^n)^{1/(n+1)}} \\ &= -\left(\frac{1}{4n^n\nu^2}\right)^{1/(n+1)} < 0\end{aligned}$$

An increase in ν implies $\mathcal{D}$ has a higher relative cost-adjusted value of successfully defending her network. Therefore, $\mathcal{D}$ will spend more resources on security, while $\mathcal{A}$ prefers not to waste resources on a prize that is relatively less valuable for her. However, as ν decreases, $\mathcal{A}$ is more willing to allocate attack resources to gain access to the high-value node. This results in $\mathcal{D}$ only increasing the accessibility of the high-value node (i.e., the number of nodes connected to the high-value node) for large values of β or n. The intuition behind this is that $\mathcal{D}$ observes a higher probability of a breach in her network when it is larger, yet she only stands to benefit from increasing access if the gains offered by a few connections is substantial, or if there are many beneficial connections that can all be protected uniformly by a single-cost security infrastructure technology.

As n gets large, $\bar{\beta}$ converges to zero, so there exists some n large enough where $\beta > \lim_{n\to\infty} \bar{\beta} = 0$:

$$\begin{aligned}\lim_{n\to\infty} \bar{\beta} &= \lim_{n\to\infty} \frac{1}{n-1}\left(\frac{n+1}{(4n^n\nu^{n-1})^{1/(n+1)}} - 1\right) \\ &= \lim_{n\to\infty} \frac{n+1}{n-1}\left(\frac{1}{4^{1/(n+1)}n^{n/(n+1)}\nu^{(n-1)/(n+1)}}\right) \\ &= \frac{1}{\nu}\lim_{n\to\infty}\frac{1}{n} = 0\end{aligned}$$

This means that $\mathcal{D}$ will provide complete access to the high-value node when she has a sufficiently large enough number of nodes to connect. However, if $\mathcal{D}$ does not have a large n (i.e., $\mathcal{D}$ is in charge of security at a small business), this may be a more difficult decision, as Figure 1.4 shows $\bar{\beta}$ is non-monotonic in n for sufficiently small values of ν. This means that when $\mathcal{A}$ has a much larger cost-adjusted value of the high-value node relative to $\mathcal{D}$, $\mathcal{D}$ will not be willing to moderately increase access unless the benefit from connecting low-value nodes to the high-value node is substantial.

1.3.2 Individual Node Defense

Equilibrium Allocations

I now look at the setting when $\mathcal{D}$ allocates defensive resources to each node individually rather than through a defensive infrastructure technology. This removes the third modification I originally made to the framework introduced in Clark and Konrad (2007), and allows me to directly analyze the impacts of the existence of an infrastructure technology in a lottery contest setting. More explicitly, let $\bar{d} = (d_1, \ldots, d_n)$ represent $\mathcal{D}$'s defensive allocations across all nodes in component C, where $c(d_i) = d_i$ for all i. The payoff functions then become:

$$u_{\mathcal{D}} = \prod_{i=1}^{m+1}\left(\frac{d_i}{d_i + a_i}\right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i=1}^{m+1} d_i \tag{1.4}$$

$$u_{\mathcal{A}} = \left(1 - \prod_{i=1}^{m+1} \left(\frac{d_i}{d_i + a_i}\right)\right) v_{\mathcal{A}} - \sum_{i=1}^{m+1} \alpha a_i \tag{1.5}$$

Similar to Lemma 2, the following result exists:

Lemma 4. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. If both $\mathcal{D}$ and $\mathcal{A}$ allocate resources to each node individually ($\bar{d} = (d_1, \ldots, d_n)$ and $\bar{a} = (a_1, \ldots, a_n)$), both $\mathcal{D}$ and $\mathcal{A}$ will optimally distribute resources uniformly across all nodes in the component containing the high-value node ($d_i^* = d_j^* = d^*$ and $a_i^* = a_j^* = a^* \ \forall \ i, j \in C$).*

Proof. Proof in Appendix A. □

This follows the same logic as before – there are no asymmetries across the nodes in C, and $\mathcal{A}$ only requires a successful attack on one node to win, so the marginal benefit from allocating resources (defensive or offensive) to a given node is equivalent across the entire network. If $\mathcal{D}$ does not allocate optimally and instead has more resources defending certain nodes, then $\mathcal{A}$ could improve her likelihood of successfully breaching the network by shifting all of her resources to the weakly defended nodes.

In addition to the uniform allocation lemma, there also exists a similar result to Lemma 3 in the following:

Lemma 5. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. Let both $\mathcal{D}$ and $\mathcal{A}$ allocate resources to each node individually ($\bar{d} = (d_1, \ldots, d_n)$ and $\bar{a} = (a_1, \ldots, a_n)$). $\mathcal{A}$ will not attack a node if $\mathcal{D}$ allocates sufficiently high defensive resources to that node $\left(a_i^* = 0 \ if \ d_i^* \geq \frac{v_{\mathcal{A}}}{\alpha}\right)$. $\mathcal{D}$ does not allocate defensive resources beyond the point at which $\mathcal{A}$ does not attack a node $\left(d_i^* \leq \frac{v_{\mathcal{A}}}{\alpha} \ if \ a_i^* = 0\right)$.*

Proof. Proof in Appendix A. □

The threshold of defensive resources at a given node needed to deter $\mathcal{A}$ from attacking that node is the same in Lemma 3 and Lemma 5. However, in this case, $\mathcal{D}$ must individually allocate this level of resources to each node in the high-value component if she wishes to deter $\mathcal{A}$ from attacking. As $\mathcal{D}$ no longer has an infrastructure technology, and therefore must pay a cost for resources allocated to each individual node, complete deterrence will be substantially more expensive in this setting.

Below, I provide the equilibrium allocations for $\mathcal{D}$ and $\mathcal{A}$ across all nodes in the component containing the high-value node. Again, the first equilibrium allocation corresponds to a state where $\mathcal{D}$ provides high security at each node and $\mathcal{A}$ is deterred from attacking. The second equilibrium allocation corresponds to the state where both $\mathcal{D}$ and $\mathcal{A}$ provide a strictly positive level of resources to the defense and attack, respectively, of the network.

Proposition 3. *Consider a model of sequential network defense between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$. $\mathcal{D}$ connects $m \in \{0, n-1\}$ low-value nodes to a high-value node and allocates defensive resources $\bar{d} = (d_1, \ldots, d_n)$ to each node. $\mathcal{A}$ observes $\mathcal{D}$'s choices and allocates attack resources $\bar{a} = (a_1, \ldots, a_n)$ to each node. $\mathcal{D}$ has a weakest-link objective, receives $v_{\mathcal{D}}(1 + m\beta)$ from a successful defense, and has linear costs. $\mathcal{A}$ has a best-shot objective, receives $v_{\mathcal{A}}$ from a successful attack, and has linear costs with the relative price of attack resources to defense resources represented by α. The outcome at each node is determined by a lottery contest success function. Given the exogeneous parameters, there exist two types of subgame perfect equilibrium allocations that can arise:* Attack *or* Deterrence.

1. **Attack:** *$\mathcal{D}$ and $\mathcal{A}$ each allocate a uniform and positive level of resources to nodes in the component containing the high-value node ($d_i^* = d^* \ \forall \ i \in C$ and $a_i^* = a^* \ \forall \ i \in C$). If $\nu < \dfrac{m+2}{1+m\beta}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, ($d^* \ \forall \ i$, $a^* \ \forall \ i$),*

are:

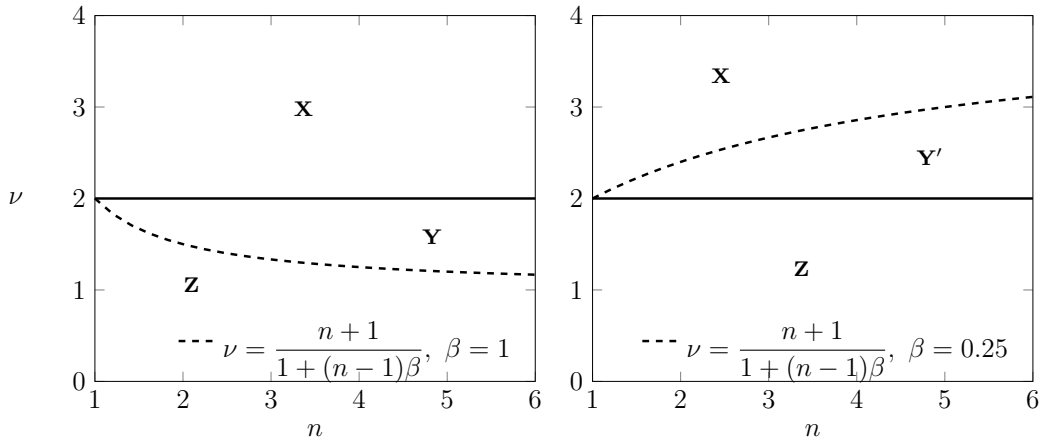
$$d^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2} \right)^{m+2}, \text{ and}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(1+m\beta)} - 1 \right]$$

2. **Deterrence:** *$\mathcal{D}$ allocates a sufficiently high level of defensive resources to each node in the component containing the high-value node such that $\mathcal{A}$ does not allocate resources to attacking $\mathcal{D}$'s network*
($a_i^* = a^* = 0 \ \forall \ i$). If $\nu \geq \dfrac{m+2}{1+m\beta}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, ($d^* \ \forall \ i$, $a^* \ \forall \ i$), are:

$$d^* = \frac{v_{\mathcal{A}}}{\alpha}, \text{ and}$$
$$a^* = 0$$

Proof. Proof in Appendix A. □

Figure 1.5 shows the regions corresponding to both types of equilibrium. Note that $\dfrac{n+1}{1+(n-1)\beta} = 2$ when $\beta = 0.5$. Therefore, region Y only exists for $\beta > 0.5$ and region Y′ only exists for $\beta \in (0, 0.5)$. It is intuitive that for small β, $\mathcal{D}$ has a higher relative valuation from isolating the high-value node than from connecting the network. As a result, she will continue to fully defend the isolated high-value node even after she stops fully defending the connected network. The converse is true when β is large, and the individual node defense setting with $\beta > 0.5$ is similar to the equilibria present in the infrastructure defense setting in Section 3.1.



(a) $\beta > 0.5$ (b) $\beta \in (0, 0.5)$

X – Deterrence for $m \in \{0, n-1\}$
Y – Deterrence for $m = n-1$, Attack for $m = 0$
Y′ – Deterrence for $m = 0$, Attack for $m = n-1$
Z – Attack for $m \in \{0, n-1\}$

Figure 1.5: Individual Node Defense: Equilibrium Regions

Individual node defense introduces region Y′, which did not exist in the infrastructure defense setting. β is small in this region, so $\mathcal{D}$ receives little benefit from connecting low-value nodes to the high-value node. As a result, isolating the high-value node offers a relatively higher marginal payoff than connecting, and the high value of ν means $\mathcal{D}$ is willing to fully protect this node when it is isolated. However, because the connected network is considerably riskier without offering much additional value compared to the isolated network, $\mathcal{D}$ will not allocate enough resources to deter $\mathcal{A}$ from attacking when the network is connected.

Another interesting element of the equilibrium settings is that only regions X and Z exist when $\beta = 0.5$. That is, when each additional connection provides $\mathcal{D}$ with a value equivalent to half the intrinsic value of the high-value node, $\mathcal{D}$ will provide (less than) full defense regardless of network size when her cost-adjusted value is at least twice (less than twice) that of $\mathcal{A}$.

Proposition 4. *Based on the subgame perfect equilibrium allocations in Proposition 3, where*

$\mathcal{D}$ allocates defensive resources to each node individually, $\bar{d} = (d_1, \ldots, d_n)$, the subgame equilibrium profits for $\mathcal{D}$ and $\mathcal{A}$, respectively, are

$$u^*_{\mathcal{D}} = \begin{cases} v_{\mathcal{D}}(1+m\beta) - (m+1)\dfrac{v_{\mathcal{A}}}{\alpha} & if\ d^* = \dfrac{v_{\mathcal{A}}}{\alpha} \\ \nu^{m+1} v_{\mathcal{D}} \left(\dfrac{1+m\beta}{m+2}\right)^{m+2} & if\ d^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

$$u^*_{\mathcal{A}} = \begin{cases} 0 & if\ d^* = \dfrac{v_{\mathcal{A}}}{\alpha} \\ v_{\mathcal{A}} \left[1 - \nu^m \left(\dfrac{1+m\beta}{m+2}\right)^{m+1} \left(\nu + m + 1 - (m+1)\nu^2 \left(\dfrac{1+m\beta}{m+2}\right)\right)\right] & if\ d^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

where $d^* = \dfrac{v_{\mathcal{A}}}{\alpha}\ \forall\ i \in C$ in the "Deterrence" subgame equilibrium and $d^* \in \left(0, \dfrac{v_{\mathcal{A}}}{\alpha}\right)\ \forall\ i \in C$ in the "Attack" subgame equilibrium.

Proof. Proof in Appendix A. □

A major difference in these equilibrium payoffs compared to the infrastructure defense setting involves the "Deterrence" equilibrium ($d^* = v_{\mathcal{A}}/\alpha$). Here, it is not immediately clear in the "Deterrence" equilibrium that $u_{\mathcal{D}}$ is increasing in the size of the network. In fact, if $v_{\mathcal{A}}/\alpha > v_{\mathcal{D}}\beta$, $\mathcal{D}$ receives a lower payoff for a larger m. Below, I provide further analysis of $\mathcal{D}$'s connection decision when nodes are defended individually.

Connection Decision

$\mathcal{D}$ initially decides between connecting $n-1$ nodes to a high-value node or isolating the high-value node. That is, $\mathcal{D}$ decides whether to defend a network of size n or of size 1. $\mathcal{D}$ should connect if $u^*_{\mathcal{D}}|_{m=n-1} \geq u^*_{\mathcal{D}}|_{m=0}$. This decision is assessed for each of the regions from Figure 1.5.

Region X:

By Proposition 3, region X corresponds to $\mathcal{A}$ being deterred from attacking either the

connected or isolated network. That is, the equilibrium allocations for $m \in \{0, n-1\}$ are:

$$(d^* \ \forall \ i, \ a^* \ \forall \ i) = \left(\frac{v_{\mathcal{A}}}{\alpha}, \ 0\right)$$

Therefore, $\mathcal{D}$ should connect her network when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{X}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{X}}|_{m=0} \\ v_{\mathcal{D}}(1+(n-1)\beta)) - n\frac{v_{\mathcal{A}}}{\alpha} &\geq v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} \\ \nu(1+(n-1)\beta) - n &\geq \nu - 1 \\ (n-1)\beta &\geq 1 + \frac{n-1}{\nu} - 1 \\ \beta &\geq \frac{1}{\nu} \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{\nu}$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. The first point to notice is that if $\beta \geq 0.5$, $\mathcal{D}$ will connect for all $\nu \geq 2$. In words, this means that $\mathcal{D}$ will increase the size of her network if the additional benefit offered from a connection is at least half the intrinsic value of the high-value node and if her relative cost-adjusted value is at least twice that of $\mathcal{A}$. If $\beta \in (0, 2)$, however, the threshold $\bar{\beta}$ (the minimum value offered from each additional connection) for $\mathcal{D}$ is inversely related to ν (the ratio of cost-adjusted values of the high-value node). This means that $\mathcal{D}$ is more likely to increase the size of her network when $\mathcal{A}$ has a relatively lower prize from a successful attack and is less likely to expend a large amount of attack resources.

Region Y:

By Proposition 3, region Y corresponds to $\mathcal{D}$ spending sufficient resources to deter $\mathcal{A}$ from attacking only when $\mathcal{D}$ connects her network. If $\mathcal{D}$ isolates the valuable node, however, $\mathcal{A}$

will attack the isolated node. This means the equilibrium allocations are:

$$(d^* \ \forall \ i, \ a^* \ \forall \ i) = \begin{cases} \left(\frac{\nu v_{\mathcal{D}}}{4}, \ \frac{(2-\nu)v_{\mathcal{D}}}{4}\right) & \text{for } m = 0 \\ \left(\frac{v_{\mathcal{A}}}{\alpha}, \ 0\right) & \text{for } m = n-1 \end{cases}$$

Using these equilibrium allocations, I find that $\mathcal{D}$ connects her network in region Y when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{Y}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{Y}}|_{m=0} \\ v_{\mathcal{D}}(1+(n-1)\beta)) - n\frac{v_{\mathcal{A}}}{\alpha} &\geq \frac{\nu v_{\mathcal{D}}}{4} \\ \nu(1+(n-1)\beta) &\geq \frac{\nu^2}{4} + n \\ (n-1)\beta &\geq \frac{\nu^2}{4\nu} + \frac{n}{\nu} - 1 \\ \beta &\geq \frac{1}{n-1}\left(\frac{\nu^2+4n-4\nu}{4\nu}\right) \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{n-1}\left(\frac{\nu^2+4n-4\nu}{4\nu}\right)$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. Region Y only exists for $\beta > 0.5$ as the upper and lower bound of the region are equivalent at $\beta = 0.5$:

$$\frac{n+1}{1+(n-1)0.5} = \frac{n+1}{0.5(n+1)} = 2$$

Furthermore, $\nu < 2$ in region Z, which ensures that $\bar{\beta}$ is strictly positive. Figure 1.6 shows the connection threshold for varying values of $\nu \in (0, 2)$:

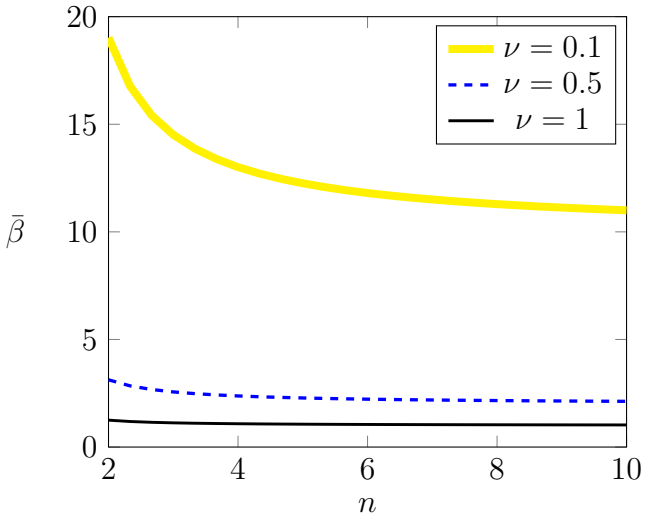


Figure 1.6: Individual Node Defense: $\mathcal{D}$'s Connection Threshold in Region Y

Here again is a common result for the connection threshold, $\bar{\beta}$, in region Y – as ν increases, $\bar{\beta}$ decreases and $\mathcal{D}$ is more likely to connect her network. This can easily be shown for this region, keeping in mind that $\nu \in (0, 2)$:

$$\begin{aligned}\frac{\partial \bar{\beta}}{\partial \nu} &= \frac{1}{n-1}\left(\frac{8\nu^2 - 16\nu - 4(\nu^2 - 4\nu + 4n)}{16\nu^2}\right) \\ &= \frac{1}{n-1}\left(\frac{\nu^2 - 4n}{4\nu^2}\right) < 0\end{aligned}$$

However, the more interesting result in this region comes from examining the connection threshold when n is large:

$$\lim_{n\to\infty} \bar{\beta} = \lim_{n\to\infty} \frac{1}{n-1}\left(\frac{\nu^2 + 4n - 4\nu}{4\nu}\right) = \frac{1}{\nu}$$

Because $\nu < 2$ in region Y, it follows that $\bar{\beta} \geq 0.5$ for all network sizes. Therefore, if the benefit to $\mathcal{D}$ from each additional connection is small (specifically less than half the intrinsic value of the high-value node), $\mathcal{D}$ will always prefer to isolate the high-value node. For larger values of β, though, $\mathcal{D}$ will only increase the size of the network if she has a sufficiently high

relative cost-adjusted value of the high-value node. Furthermore, if β is restricted to the range $(0, 1)$ – not an unrealistic assumption if one believes the high-value node has more value than what is offered from each individual connection – then $\mathcal{D}$ will never connect her network if $\mathcal{A}$ has a greater relative cost-adjusted value from winning the contest.

Region Y′:

By Proposition 3, region Y′ corresponds to $\mathcal{D}$ spending sufficient resources to deter $\mathcal{A}$ from attacking only when $\mathcal{D}$ chooses to isolate the high-value node. If $\mathcal{D}$ connects her network, however, $\mathcal{A}$ will attack the network. This means the equilibrium allocations are:

$$(d^* \,\forall\, i,\ a^* \,\forall\, i) = \begin{cases} \left(\dfrac{v_{\mathcal{A}}}{\alpha},\ 0\right) & \text{for } m = 0 \\ \left(\nu^n v_{\mathcal{D}} \left(\dfrac{1+(n-1)\beta}{n+1}\right)^{n+1},\ d^*\left[\dfrac{n+1}{\nu(1+(n-1)\beta)} - 1\right]\right) & \text{for } m = n-1 \end{cases}$$

Using these equilibrium allocations, I find that $\mathcal{D}$ connects her network in region Y′ when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{Y}'}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{Y}'}|_{m=0} \\ \nu^n v_{\mathcal{D}} \left(\frac{1+(n-1)\beta}{n+1}\right)^{n+1} &\geq v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} \\ \left(\frac{\nu(1+(n-1)\beta)}{n+1}\right)^{n+1} &\geq \nu - 1 \\ 1 + (n-1)\beta &\geq \frac{(n+1)(\nu-1)^{1/(n+1)}}{\nu} \\ \beta &\geq \frac{1}{n-1}\left(\frac{(n+1)(\nu-1)^{1/(n+1)}}{\nu} - 1\right) \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{n-1}\left(\frac{(n+1)(\nu-1)^{1/(n+1)}}{\nu} - 1\right)$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. Note that $\nu \geq 2$ in region Y′, so $\nu - 1 > 0$. Furthermore, region Y′ only exists for $\beta \in (0, 0.5)$, as $\beta > 0$ by definition, and the upper and lower bounds

of the region are equivalent at $\beta = 0.5$:

$$\frac{n+1}{1+(n-1)0.5} = \frac{n+1}{0.5(n+1)} = 2$$

Figure 1.7 shows the threshold, $\bar{\beta}$, for various values of $\nu > 2$.

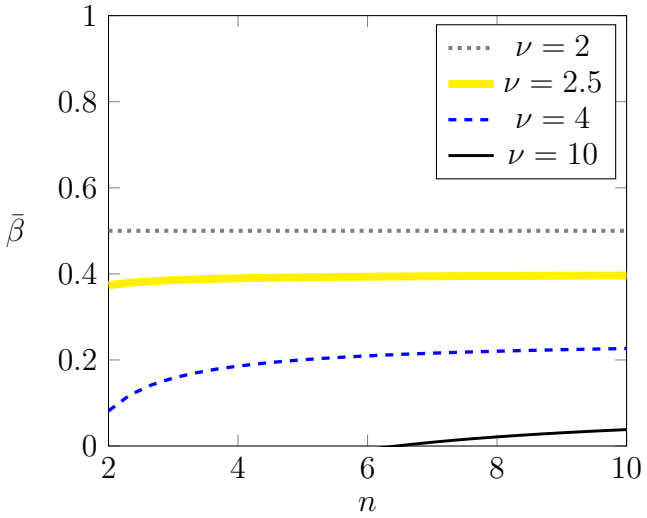


Figure 1.7: Individual Node Defense: $\mathcal{D}$'s Connection Threshold in Region Y$'$

The most immediate result is that $\mathcal{D}$ will not connect if $\nu = 2$ in this region. This is because $\beta \in (0, 0.5)$ must be true for region Y$'$ to exist. However, if $\nu = 2$, then $\bar{\beta} = 0.5 > \beta$ for any β in region Y$'$. Another important result here is that the minimum value offered from each additional connection, $\bar{\beta}$, is decreasing as the ratio of cost-adjusted values of the high-value node, ν, increases. This implies that $\mathcal{D}$ is more likely to connect for a higher value of ν, which is consistent with the result found in region X. This can be shown by the following

comparative static:

$$\begin{aligned}
\frac{\partial \bar{\beta}}{\partial \nu} &= \frac{n+1}{n-1}\left(\frac{\nu\left(\frac{1}{n+1}\right)(\nu-1)^{-n/(n+1)} - (\nu-1)^{1/(n+1)}}{\nu^2}\right) \\
&= \frac{n+1}{n-1}\left(\frac{\nu}{\nu^2(n+1)(\nu-1)^{n/(n+1)}} - \frac{(\nu-1)^{1/(n+1)}}{\nu^2}\right) \\
&= \frac{1}{n-1}\left(\frac{\nu-(n+1)(\nu-1)}{\nu^2(\nu-1)^{n/(n+1)}}\right) \\
&= \frac{1}{n-1}\left(\frac{n(1-\nu)+1}{\nu^2(\nu-1)^{n/(n+1)}}\right) < 0
\end{aligned}$$

The final point to make about this region is that as n gets large, there exists a strictly positive minimum value of β needed for $\mathcal{D}$ to connect.

$$\begin{aligned}
lim_{n\to\infty}\bar{\beta} &= \lim_{n\to\infty}\left(\frac{n+1}{n-1}\left(\frac{(\nu-1)^{1/(n+1)}}{\nu}\right) - \frac{1}{n-1}\right) \\
&= \frac{1}{\nu}\lim_{n\to\infty}\frac{n+1}{n-1} \\
&= \frac{1}{\nu}
\end{aligned}$$

This shows that $\mathcal{D}$'s decision to connect is equivalent between regions X and Y$'$ for large values of n. That is, $\mathcal{D}$ is more likely to connect as their cost-adjusted value of the high-value node increases relative to $\mathcal{A}$.

Region Z: By Proposition 3, $\mathcal{A}$ allocates a strictly positive level of resources to attack either a connected or isolated network in region Z. This means the equilibrium allocations for $m \in \{0, n-1\}$ are:

$$\begin{aligned}
d^* &= \nu^{m+1} v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2} \quad \forall\ i \\
a^* &= \nu^{m+1} v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2}\left[\frac{m+2}{\nu(1+m\beta)} - 1\right] \quad \forall\ i
\end{aligned}$$

Using these equilibrium allocations, I find that $\mathcal{D}$ connects her network in region Z when:

$$\begin{aligned} u^*_{\mathcal{D},\mathrm{Z}}|_{m=n-1} &\geq u^*_{\mathcal{D},\mathrm{Z}}|_{m=0} \\ \nu^n v_{\mathcal{D}} \left(\frac{1+(n-1)\beta}{n+1}\right)^{n+1} &\geq \frac{\nu v_{\mathcal{D}}}{4} \\ 1+(n-1)\beta &\geq \left(\frac{1}{4\nu^{n-1}}\right)^{1/(n+1)} \\ \beta &\geq \frac{1}{n-1}\left(\left(\frac{1}{4\nu^{n-1}}\right)^{1/(n+1)} - 1\right) \end{aligned}$$

This implies the threshold, $\bar{\beta}$:

$$\bar{\beta} = \frac{1}{n-1}\left(\left(\frac{1}{4\nu^{n-1}}\right)^{1/(n+1)} - 1\right)$$

such that $\mathcal{D}$ connects if $\beta \geq \bar{\beta}$. Note that $\nu \in (0, 2)$ in region Z. This connection threshold is expressed graphically for various values of ν in Figure 1.8. For very small values of ν and small n, the connection threshold for $\mathcal{D}$ will be great. However, as the network size gets large, the threshold $\bar{\beta}$ converges toward zero.

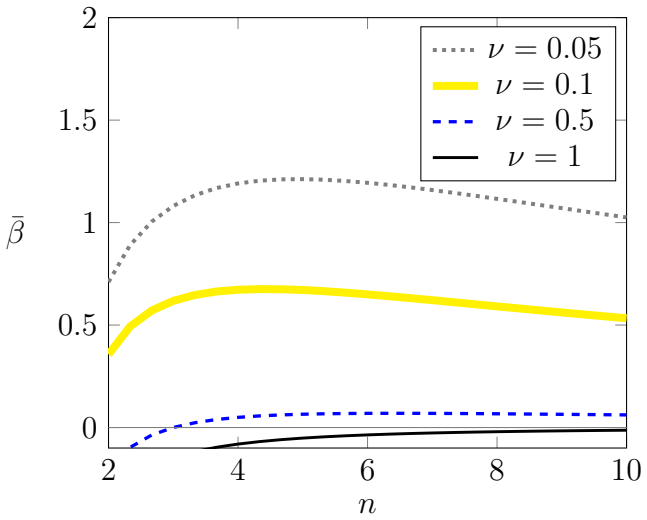


Figure 1.8: Individual Node Defense: $\mathcal{D}$'s Connection Threshold in Region Z

Similar to previous analyses, $\bar{\beta}$ is again decreasing in ν in region Z:

$$\frac{\partial \bar{\beta}}{\partial \nu} = \frac{1}{(n+1)4^{1/(n+1)}} \left(-\frac{1}{\nu^{2n/(n+1)}} \right) < 0$$

Therefore, $\mathcal{D}$ is more likely to connect her network when she has a higher relative cost-adjusted value of the high-value node. The intuition for this is that $\mathcal{A}$ will expend few resources attacking the network when she receives a relatively low value from a successful attack, so $\mathcal{D}$ can risk the vulnerability that comes with a larger network.

Similarly to region Z in Section 3.1, as n gets large, $\bar{\beta}$ converges to zero, so there exists some n large enough where $\beta > \lim_{n\to\infty} \bar{\beta} = 0$:

$$\begin{aligned}
\lim_{n\to\infty} \bar{\beta} &= \lim_{n\to\infty} \frac{1}{n-1} \left(\left(\frac{1}{4\nu^{n-1}} \right)^{1/(n+1)} - 1 \right) \\
&= \lim_{n\to\infty} \frac{1}{n-1} \left(\frac{1}{4^{1/(n+1)} \nu^{(n-1)/(n+1)}} \right) \\
&= \frac{1}{\nu} \lim_{n\to\infty} \frac{1}{n-1} = 0
\end{aligned}$$

Therefore, at a sufficiently large n, $\mathcal{D}$ will provide complete access to the high-value node. However, for small values ν, $\bar{\beta}$ may be large, and increasing, for small networks. This threshold will eventually begin decreasing at some larger value of n. When $\mathcal{A}$ has a far greater cost-adjusted value of the high-value node relative to $\mathcal{D}$ (small ν), though, $\mathcal{D}$ will prefer isolating the high-value node to adding a minor level of access within her network.

1.4 Conclusion

I studied a sequential weakest-link network defense contest where the defender has a choice of how many nodes to defend. The defender receives additional benefit for each node in the network; however, with more nodes, the weakest-link network is more vulnerable to attack.

After the defender decides on the size of the network and allocates defensive resources, the attacker observes the defender's choices and allocates attack resources (or does not attack). Each node is successfully defended with a probability proportional to the resource allocations of each player. The defender wins the contest if she defends every node in the network (weakest-link objective), and the attacker wins the contest if she successfully attacks at least one node in the network (best-shot objective). I consider two settings: (1) the defender allocates resources to an infrastructure technology which uniformly protects all nodes in the network, and (2) the defender allocates defensive resources to each node individually. Under both settings, I examine a defender's decision to increase the size of their network under threat from an adversary.

My primary results can be summarized in three main points. First, as the defender's cost-adjusted value of an important asset increases relative to the attacker's value, the minimum marginal benefit of an additional connection needed for the defender to allow access to this asset decreases. For example, suppose the defender in this model can remove personal identifiers from a database to make it less valuable to potential attackers, then the defender will be able to allow greater access to this database (i.e., more employees or contracting agents working with the data) without a substantial increase in the network's vulnerability (as the attacker is allocating less resources to individual nodes due to a lower valuation of the asset). This result is ubiquitous across all equilibrium regions and both infrastructure and individual node defense settings.

The second primary result is that the minimum marginal benefit of an additional connection needed for the defender to connect their network generally decreases as the potential size of the network increases. That is, the minimum marginal benefit of an additional connection needed for the defender to connect is generally smaller when the defender is deciding to connect twenty additional nodes than when the defender is deciding to connect ten additional nodes. However, there are a few exceptions to this general rule in specific equilibrium regions.

The final primary result involves the distinction between infrastructure defense and individual node defense. Due to the nature of infrastructure defense, it is not costly for the defender to increase the size of the network she is protecting. This means that the defender will be able to deter attacks completely when she has a relatively high cost-adjusted valuation of the high-value node, and as a result, she will fully connect her network. The same result does not apply in the individual node defense setting, as the defender may prefer isolating the high-value node even when she would allocate enough resources to deter attacks regardless of network size. Another distinction between these settings comes when the defender has only a slightly larger cost-adjusted valuation of the high-value node than the attacker. Under these circumstances, the defender will always connect a large network in the infrastructure defense case, but will not in the individual node defense case unless the marginal benefit of each additional connection is large. Future research should explore the limitations of this result when relaxing the assumptions made by this model.

While this paper provides an initial analysis of a network defender's decision between efficiency and security, more work needs to be done to explore this decision. For example, I consider a simple choice for the defender to either defend an isolated network or a completely connected network. The next step in this line of research would be to specifically examine the optimal network size which maximizes the defender's payoff. Furthermore, my analysis introduces infrastructure defense in a lottery contest setting, and I compare results from this setting to an individual node defense setting. More should be done, however, to explore a setting that allows for a mixture of infrastructure and individual node defense, as a mixture of these two defensive systems is much more applicable to real-world security systems. A final potential direction to extend the research within this paper is through the use of different functions to represent the valuation of additional connections within the network. I use a simple linear model to keep the mathematical analysis simple, though exploring different functional forms would likely yield extremely interesting and practical results.

Chapter 2

An Experimental Study of Conjectural Equilibrium: Limited Feedback in a Threshold Public Good Game

Co-Authored with Michael McBride

2.1 Introduction

In many settings, individuals receive limited feedback about the actions of others. For example, individuals learn about a macro outcome such as a price without observing the micro decisions that generate that price, or members of a team learn whether their team succeeded without observing the individual members' contributions made on behalf of the team. Because only partial information about others' actions is obtained in such settings, an individual may form a belief that is correct about the macro outcome but not the micro behaviors, and those incorrect beliefs may persist over time.

The *Conjectural Equilibrium* concept (Gilli, 1999) was created to study settings with limited feedback—what is called *imperfect monitoring* in the game-theoretic literature. In the standard *Nash Equilibrium*, each individual plays a best response to the strategies of the other players so that each individual is implicitly assumed to have correct beliefs about other players' strategies. However, in a Conjectural Equilibrium, each individual chooses a best response to their beliefs about the other players' strategies, and their beliefs can be incorrect as long as their beliefs are not contradicted by the limited feedback received. By relaxing Nash Equilibrium's correct-beliefs assumption, the Conjectural Equilibrium concept thus identifies a richer set of strategies that may emerge in the long run in limited-feedback settings.

Though limited feedback is a salient feature of many real-world settings and many experimental studies, the predictive accuracy of the Conjectural Equilibrium concept has not been rigorously tested in experiments. This paper helps to fill this gap in the literature by experimentally testing the merits of the Conjectural Equilibrium concept in a limited-feedback setting. Our main question is: Does the Conjectural Equilibrium concept provide better predictions of behavior than the Nash Equilibrium concept when individuals have limited feedback? However, we also seek to examine key assumptions behind the Conjectural Equilibrium concept to identify why it does or does not outperform Nash Equilibrium. Specifically: Do subjects persist in holding incorrect beliefs under limited feedback? Do players get stuck in non-Nash behaviors under limited feedback? And do players choose best responses at similar rates across feedback settings?

For our setting, we consider a threshold public good (TPG) game in which a public good is provided only if the combined contributions of the group exceed a known threshold. TPG games are a natural setting in which to test Conjectural Equilibrium predictions for two reasons. First, many real-world settings with imperfect feedback have features of TPG games. For example, many group interactions have a well-defined sense of success (e.g., a

group of students who want a passing grade on a group assignment or a team of attorneys who want to win a trial) while also having limited feedback about the contributions by group members. Second, the set of Conjectural Equilibria in a standard TPG game will differ in concrete ways from the set of Nash Equilibria as feedback becomes more and more limited. We can thus obtain clear predictions about how the set of Conjectural Equilibria differs from the set of Nash Equilibria as the amount of feedback changes, thereby allowing for a test of usefulness of the Conjectural Equilibrium concept.

Our 3×2-factorial experiment varies the level of feedback and the value of the public good. Feedback is provided at one of three levels: *Full* (feedback about the group outcome and individual actions), *Partial* (feedback about the group outcome but not individual actions), and *None* (no feedback about the group outcome or individual actions). The public good values takes one of two values: *High* and *Low*. In our hybrid design, the subjects in each experimental session keep the same public good value for forty rounds. Half of the subjects play twenty rounds under None and then twenty rounds under Partial, and the other half play twenty rounds under None and then twenty rounds under Full.

In addition to making a decision to contribute in each round, subjects report beliefs about others' actions in each round and undertake separate tasks at the end of the experimental session to measure risk aversion, other-regardingness, and cognitive reflection. These data allow us to examine why Conjectural Equilibrium does or does not perform well as a predictive concept. One possibility is that subjects have idiosyncratic traits that lead to differences in behavior across the feedback conditions beyond what is captured by a standard model. Subjects might also play dynamic (farsighted) strategies so that their decisions in any particular round do not maximize their single-period expected payoffs, and these strategies may lead to differences in stability across different feedback conditions.

Overall, we find moderate support for the Conjectural Equilibrium concept's predictive power. Strategy profiles that are not Nash Equilibria but are Conjectural Equilibria do

generally increase in frequency as feedback becomes more limited. Moreover, beliefs are less accurate as feedback becomes more limited, subjects are less likely to break out of non-Nash Conjectural Equilibria under limited feedback, and other-regarding preferences do not interact with changes in feedback. These findings provide experimental evidence that justifies the key conceptual motivations behind using Conjectural Equilibrium instead of Nash Equilibrium in limited-feedback settings, namely, incorrect equilibrium beliefs due to limited feedback and the subsequent effects on behavior.

However, other findings place a limit on Conjectural Equilibrium's usefulness. For one, the observed strategy profiles best match our hypotheses when controlling for subjects' risk aversion, indicating that risk aversion plays just as strong a role in affecting behavior as does limited feedback. More problematic is that subjects' risk aversion interacts with feedback in a way not accounted for by the theory, which makes precise Conjectural Equilibrium predictions difficult to obtain. Convergence in beliefs is also slower under limited feedback when the public good value is Low, so Conjectural Equilibrium's predictions are better tested when actors have relatively long time horizons to reach an equilibrium. Future theoretical and empirical studies should take these findings into account.

Our paper contributes to two specific literatures. The first is the game-theoretical literature on non-Nash Equilibrium concepts, e.g., Battigalli et al. (1992); Rubinstein and Wolinsky (1994); Gilli (1999); Dekel et al. (1999); Azrieli (2009); Esponda (2013); Battigalli et al. (2015).[1] This literature examines how incorrect beliefs can persist in limited-feedback settings and proposes equilibrium concepts that are better suited to those contexts than Nash Equilibrium. Among these non-Nash concepts is Conjectural Equilibrium which allows players to have incorrect beliefs about actions on the equilibrium path, a plausible possibility under imperfect monitoring (Gilli, 1999). Our paper provides the first experimental test of

[1]The Conjectural Equilibrium has also been used outside of economics to study limited-feedback settings, e.g., Wellman and Hu (1998) and Kalashnykova et al. (2021). It has also inspired new equilibrium concepts in network formation games, e.g., McBride (2006b), McBride (2006c), and McBride (2008).

the Conjectural Equilibrium concept in any simple economic game.

The second literature is on TPG games (also called "discrete public goods" or "step-level public goods"). Palfrey and Rosenthal (1984) provided the first game-theoretic examination of TPG games, and their seminal work has sparked a large body of theoretical and experimental research. Some theoretical studies include Nitzan and Romano (1990), Bagnoli and Lipman (1992), Suleiman (1997), Menezes et al. (2001), and McBride (2006a); some experimental studies include Offerman et al. (1996), Marks and Croson (1998), Wit and Wilke (1999), Offerman et al. (2001), Au (2004), and McBride (2010). Our study is the first theoretical and experimental examination of TPG games using the Conjectural Equilibrium concept. Our theoretical analysis shows how the Conjectural Equilibrium concept can be applied to TPG games with different levels of feedback, and our experimental results show that the Conjectural Equilibrium provides mixed predictive power for those limited-feedback TPG games.

Finally, our paper also contributes to the broader experimental economics literature as a whole. Deciding how much feedback to provide subjects is a task for any experimental economist in any experimental study, yet this decision is often made in accordance with common sense and experience rather than explicit guidance from an equilibrium concept. Our paper demonstrates not only how the Conjectural Equilibrium concept can be used to generate predictions, but also that it can have predictive power

2.2 Theory

2.2.1 Imperfect Monitoring

In game-theoretic notation, a standard normal-form game G is a combination

$$G = \langle I, \{S_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle, \tag{2.1}$$

where $I = \{1, 2, 3, ..., n\}$ is the set of players, S_i is the set of strategies for player $i \in I$, and u_i is the utility function for player $i \in I$.

We add two additional objects to account for imperfect monitoring (limited feedback). First, we denote each player's *beliefs*. Let $\pi_i \in \triangle(S)$ be a probability distribution over the set of strategy profiles that represents i's beliefs over the strategies being played. Second, we denote each player's feedback in the form of an *information partition* P_i over the set of strategy profiles $S = \times_{i \in I} S_i$. Let $P_i(s) \subseteq S$ denote the part of the partition that contains s. Thus, for all i: (i) $P_i(s) = P_i(s')$ for any s and s', $s \neq s'$, that are in the same part, (ii) $P_i(s) \cap P_i(s') = \varnothing$ for any s and s', $s \neq s'$, that are not in the same part, and (iii) $\cup_s P_i(s) = S$.

The interpretation of the partition is that if $s \in P_i(s)$ for player i, then, when the true profile of strategies played by the players is s, player i cannot distinguish whether the true state is s or any other $s' \in P_i(s)$. That is, player i knows that the true state is one of the strategy profiles in $P_i(s)$ but cannot distinguish which state in $P_i(s)$ is the true state.

A game with perfect monitoring must have perfectly discriminating information partitions, i.e., for every player $i \in I$, it must be that $P_i(s) = s$ for each $s \in S$ so that $P_i(s) \cap P_i(s') = \varnothing$ when $s \neq s'$. Conversely, a game with imperfect monitoring must have at least one (s, s') pair, $s \neq s'$, with $s, s' \in P_i(s)$ for at least one player $i \in I$. In this scenario, player i cannot

always discern the actions of another player.

We can now define a *game of imperfect monitoring* G_{IM} to be a combination

$$G_{IM} = \langle I, \{S_i\}_{i \in I}, \{u_i\}_{i \in I}, \{P_i\}_{i \in I}, \{\pi_i\}_{i \in I} \rangle. \tag{2.2}$$

This definition of a game is identical to that of a standard game but with the addition of the information partition and the beliefs. As we see next, these additions enable a formal analysis of incorrect beliefs in equilibrium.

2.2.2 Conjectural Equilibrium

Letting $b_i : \triangle S \to S_i$ denote i's (pure) best response function, a (pure) Nash Equilibrium of game G is a strategy profile $s^* = (s_1^*, \ldots, s_n^*)$ such that $s_i^* \in b_i(s^*)$ for all $i \in I$. However, there is an equivalent definition that reveals the correct-beliefs assumption that is implicit in the Nash Equilibrium concept: A (pure) Nash Equilibrium is a combination (s_i^*, π_i^*) for each $i \in I$ such that (i) $s_i^* \in b_i(\pi_i^*)$ and

$$\text{(ii) } \pi_i^*(s') = \begin{cases} 1, & \text{if } s' = s^*, \\ 0, & \text{if } s' \neq s^*, \end{cases} \tag{2.3}$$

where $b_i(s) = \{s_i' \in S : u_i(s_i', s_{-i}) \geq u_i(s_i'', s_{-i}) \, \forall s_i'' \in S\}$ is player i's best-response correspondence.

In words, each player in a Nash Equilibrium is (i) playing a best response to their beliefs and (ii) their beliefs are correctly assigning probability 1 to the strategies actually played by the other players, i.e., each player's beliefs assign probability 1 to the "true" strategy profile.

However, one criticism of Nash Equilibrium is that the correct-beliefs assumption is too strict

in settings with imperfect monitoring. With imperfect monitoring, players might learn only partial information about the actions of other players, and if that feedback is sufficiently limited, then their beliefs should not necessarily be expected to converge to the truth. The Conjectural Equilibrium concept is argued to be more appropriate than Nash Equilibrium when there is imperfect monitoring because it allows players to hold beliefs that are incorrect as long as they are not violated by available evidence. Thus, the set of Conjectural Equilibria is the set of all possible steady states that might be reached under any learning dynamic when learning is constrained by limited feedback in the form of imperfect monitoring.

To define a Conjectural Equilibrium, condition (ii) from the Nash Equilibrium definition is modified: A (pure) Conjectural Equilibrium is a combination (s_i^*, π_i^*) for each $i \in I$ such that (i) $s_i^* \in b_i(\pi_i^*)$ and

$$\begin{aligned} &\text{(ii-a) For any } s' \text{ with } \pi_i^*(s') > 0, \text{ it must be that } s' \in P_i(s^*), \\ &\text{(ii-b) } \pi_i^*(s') = 0 \text{ for any } s' \notin P_i(s^*). \end{aligned} \tag{2.4}$$

Condition (ii-a) allows players' beliefs to have non-zero probability on a state s' that is not the true state s^* as long as s' and s^* are in the same part of the partition, i.e., $s', s^* \in P_i(s^*)$. That is, player i's beliefs can incorrectly assign positive probability to a state s' that is not the true state s as long as player i cannot distinguish state s' from state s^*. Condition (ii-b) says that a player is not allowed to hold incorrect beliefs when those beliefs are contradicted by the feedback, i.e., when the player's information partition can distinguish the state s' from the true state s^*. Keeping condition (i) unchanged means that both Nash Equilibrium and Conjectural Equilibrium assume that players select best responses to their beliefs; the only difference between the two is that Conjectural Equilibrium allows beliefs to be incorrect as long as they are not contradicted by available evidence, while Nash Equilibrium never allows beliefs to be incorrect.

There are three features of Conjectural Equilibria that follow immediately from its definition. First, every Nash Equilibrium is a Conjectural Equilibrium in which players have correct beliefs. That fact implies that there will always exist a Conjectural Equilibrium (because there always exists a Nash Equilibrium, though it might be in mixed strategies). Second, if the game has perfect monitoring, then the set of Conjectural Equilibria is identical to the set of Nash Equilibria. Intuitively, an individual's beliefs must be correct under perfect feedback because their information partition is perfectly discriminating. Third, if the game has limited feedback, then there may exist a Conjectural Equilibrium that is not a Nash Equilibrium because one or more players have incorrect beliefs that are not contradicted by evidence. It is this third feature that motivates the use of Conjectural Equilibrium as a concept for settings with limited feedback.

Because the set of Conjectural Equilibria (weakly) increases as feedback becomes increasingly limited, the equilibrium-selection problem can also worsen as feedback becomes more limited. This means that the predictive power of Conjectural Equilibrium is not an ability to predict a single particular strategy profile (i.e., a point prediction) but rather to identify a set of possible equilibrium profiles. As will be demonstrated in detail below, it is this notion of prediction that we use in this paper. Although this may seem like a flaw in the Conjectural Equilibrium concept, the equilibrium-selection problem can exist even with Nash Equilibrium, so it does not imply that the Conjectural Equilibrium concept is fundamentally different from Nash Equilibrium in this regard. At the same time, the relaxation of the correct-beliefs assumption is both realistic and practical when studying limited feedback. It is our view that the worsening of the equilibrium-selection problem is merely the cost to be paid for the added realism of allowing for incorrect beliefs in equilibrium.

2.2.3 The Threshold Public Good Game

In the standard threshold public good (TPG) game, each $i \in \{1, ..., n\}$ chooses $s_i \in \{0, 1\}$ to maximize utility

$$u_i(s) = \begin{cases} v - s_i, & \text{if } \sum_{j \in I} s_j \geq t, \\ -s_i, & \text{otherwise.} \end{cases} \tag{2.5}$$

The interpretation is that $s_i = 1$ means player i contributed to the public good at cost 1, while $s_i = 0$ means player i did not contribute. If t or more players contribute, then the public good worth value v is provided to everybody, regardless of whether or not they contributed. Otherwise, the public good is not provided, and the players do not receive a refund for their contributions. The no-refund property is appropriate for settings in which the player's contribution is an expended effort or resource, and it creates an additional strategic element to contributing as will be shown below.

Our experiment uses $n = 3$, $t = 2$, and $v > 1$, in which case the game has two types of pure Nash Equilibria (note that there are also mixed Nash Equilibria). The first is the *no-contribution* Nash Equilibrium in which $s_i = 0$ for all i. The second is a *perfect-provision* Nash Equilibrium with exactly $t = 2$ contributors.

Proposition 5. *In the threshold public good game with $n = 3$, $t = 2$, and $v > 1$:*

(a) The set of pure Nash Equilibria includes the no-contribution and perfect-provision strategy profiles.

(b) The set of pure Nash Equilibria is the same for all levels of feedback.

The proofs of both propositions are in Appendix B, but the intuition is straightforward. Perfect provision is an equilibrium because both of the two contributors are pivotal while a third contributor would be redundant. Perfect provision is also efficient and maximizes

the sum of players' utilities. There are $\binom{n}{t} = \binom{3}{2} = 3$ of these perfect-provision, pure Nash Equilibria, each with a different combination of contributors. Each perfect-provision equilibrium yields inequality in payoffs because the one non-contributor is a free rider who benefits from the two contributors' efforts without paying the cost of contributing. The no-contribution profile is also an equilibrium because contributing alone is worse than not contributing at all. The no-contribution equilibrium yields equal payoffs, but it is inefficient because each individual's payoff is strictly higher if the threshold is met. Having just a single contributor cannot be a Nash Equilibrium because the sole contributor is better off not contributing and because one of the non-contributors can become a pivotal contributor by contributing. Similarly, three contributors is not a Nash Equilibrium because any of the contributors would be better off becoming a free rider.

2.2.4 Limited Feedback in the Threshold Public Good Game

We consider three feedback settings in the TPG game. Perfect monitoring consists of perfectly-discriminating information partitions, i.e., $P_i^{PERFECT}(s) = \{s\}$ for each $s \in S$. We will consider three limited-feedback settings that progressively reduce the feedback from this perfect monitoring benchmark.

The first is what we call *Full* feedback. Our Full feedback is not perfect monitoring, but we call it Full because it is sufficient feedback so that the set of Conjectural Equilibria equals the set of Nash Equilibria. Specifically, Full feedback is when the subject knows their own contribution choice and learns the exact number of other contributors in their group but does not learn the identities of other contributors. Letting the players be denoted as i, j, and k, with $s = (s_i, s_j, s_k)$ so that i's strategy is listed first in s, then player i's Full feedback

information partition is:

$$P_i^{FULL}(s) = \begin{cases} \{(0,0,0)\}, & \text{if } s = (0,0,0), \\ \{(0,1,0),(0,0,1)\}, & \text{if } s \in \{(0,1,0),(0,0,1)\}, \\ \{(0,1,1)\}, & \text{if } s = (0,1,1), \\ \{(1,0,0)\}, & \text{if } s = (1,0,0), \\ \{(1,1,0),(1,0,1)\}, & \text{if } s \in \{(1,1,0),(1,0,1)\}, \\ \{(1,1,1)\}, & \text{if } s = (1,1,1). \end{cases} \tag{2.6}$$

Observe how the number of other contributors is the same for each strategy profile within the same part of the partition.

With *Partial* feedback, each player knows their own contribution and learns whether the good is provided or not but not the exact number of contributions. The Partial feedback information partition is

$$P_i^{PARTIAL}(s) = \begin{cases} \{(0,0,0),(0,1,0),(0,0,1)\}, & \text{if } s \in \{(0,0,0),(0,1,0),(0,0,1)\}, \\ \{(0,1,1)\}, & \text{if } s = (0,1,1), \\ \{(1,0,0)\}, & \text{if } s = (1,0,0), \\ \{(1,1,0),(1,0,1),(1,1,1)\}, & \text{if } s \in \{(1,1,0),(1,0,1),(1,1,1)\}. \end{cases} \tag{2.7}$$

Observe that the strategy profile can be perfectly discerned by i if i is the only non-contributor or if i is the only contributor, but it cannot be perfectly discerned if the good is not provided and i does not contribute or if the good is provided and i contributes.

In the *None* feedback setting, the player knows their own contribution but learns nothing about the contributions of other players and whether or not the good is provided. The None

feedback information partition is

$$P_i^{NONE}(s) = \begin{cases} \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}, & \text{if } s \in \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}, \\ \{(1,0,0),(1,1,0),(1,0,1),(1,1,1)\}, & \text{if } s \in \{(1,0,0),(1,1,0),(1,0,1),(1,1,1)\}. \end{cases} \tag{2.8}$$

Now player i can only discern whether they contribute or not, so there are only two parts of the information partition—one for when i contributes and one for when i does not contribute.

Unlike for Nash Equilibirum, the set of Conjectural Equilibria changes as feedback changes.

Proposition 6. *In the threshold public good game with $n = 3$, $t = 2$, and $v > 1$:*

(a) Under Full feedback, the set of pure Conjectural Equilibria is equivalent to the set of pure Nash Equilibria, i.e., it consists of the no-contribution and perfect-provision strategy profiles.

(b) Under Partial feedback, the set of pure Conjectural Equilibria includes the pure Nash Equilibria and the three-contributor strategy profile.

(c) Under None feedback, the set of pure Conjectural Equilibria includes the pure Nash Equilibria, the three-contributor strategy profile, and the one-contributor strategy profile.

Figure 2.1 summarizes the pure Conjectural Equilibria for the three feedback settings. Each number corresponds to the number of contributors in a strategy profile. For each feedback setting, the dark gray for 2 signifies an efficient, perfect-provision equilibrium with two contributors, and the light gray denotes an inefficient equilibrium. Under Full feedback, there is only the efficient, perfect-provision (two-contributor) equilibrium and the inefficient, no-contributor equilibrium. These are the two Nash Equilibria, which correspond to Conjectural Equilibria that happen to have correct beliefs. That 1 and 3 are not shaded in the top row indicates there is no equilibrium with one or three contributors under Full feedback.

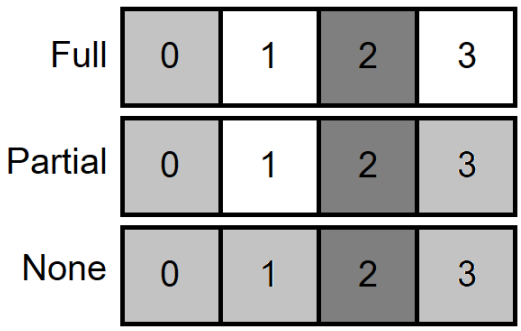


Figure 2.1: Summary of Conjectural Equilibria in the Threshold Public Good Game
Dark gray denotes efficient equilibrium, light gray denotes inefficient equilibrium, and white denotes not equilibrium.

The light gray for 3 under Partial feedback indicates that there exists a Conjectural Equilibrium where subjects have incorrect beliefs. Such can be the case, for example, if each player believes that they are one of two contributors, not realizing that there is a third contributor. However, having one contributor cannot be a Conjectural Equilibrium strategy profile because a player who contributes but learns that the public good was not provided will be able to infer that they were the only contributor, and thus know they are better off by not contributing.

Any pure strategy profile can be sustained as a Conjectural Equilibrium under None feedback. If the player believes that they are pivotal, a belief that is not violated when there is None feedback, then contributing is a best response. If the player believes that they are not pivotal, which is another belief that is not violated, then not contributing is best response.

That the set of equilibria increases as feedback decreases is a direct consequence of having more beliefs possible in equilibrium, yet the structure of the information partitions places constraints on how the set of equilibria increases. If we assume that individuals reach a Conjectural Equilibrium but do not know which Conjectural Equilibrium they will reach, we can at least say that when feedback decreases from Full to Partial, then we may expect to observe an increase in three-contributor profiles but not one-contributor profiles. Continuing the logic, we may expect to see an increase in one-contributor profiles when feedback decreases from Partial to None. The changes in the set of Conjectural Equilibria caused

by changes in feedback thus provide testable predictions to take to the laboratory.

2.3 Experimental Design

The experiment was conducted at the Experimental Social Science Laboratory (ESSL) at the University of California, Irvine. Subjects were recruited from the ESSL subject pool using ESSL's standard recruiting procedures. UC Irvine students register to be in the ESSL subject pool via an online registration system after learning of ESSL from fellow students or email announcements. Members of the ESSL Subject Pool then receive emails about upcoming experimental sessions, and they are invited to sign up for particular sessions. They can sign up for at most one session for each experimental study. Subjects who have signed up receive a reminder email before the session.

Each participant received a $10 show-up payment as well as additional earnings based on their decisions made during the experiment. Each experimental session lasted about 55 minutes, and subjects earned about $20 on average for participating in one session. All payments were made electronically using Zelle, Paypal, or Venmo.

The experiment took place remotely, with subjects entering the experiment at the scheduled time using a special URL that brought them to our oTree experimental software (Chen et al., 2016). This experiment involved no deception, and all decision tasks and payment information were clearly explained via on-screen instructions. These general procedures conform to the standard practice in experimental economics and the protocol narrative approved by our Institutional Review Board (HS #2011-8378). This study was pre-registered with the Open Science Foundation in October 2023 before we conducted our experimental sessions.[2] All experimental sessions were conducted in accordance with our pre-registration.

[2]See "Testing Conjectural Equilibrium Predictions: The Effect of Feedback in Threshold Public Good Games," DOI 10.17605/OSF.IO/PBK6W.

The experiment featured a 3×2 hybrid design with: the three levels of feedback examined above, i.e., Full, Partial, and None; and two values of the public good, i.e., High ($v = 4$) and Low ($v = 1.6$). Each subject was treated with a single value of the public good (either high or low) and two levels of feedback (either None and Partial or None and Full). We use a within-subject design for None/Partial and None/Full to conserve subjects. Additionally, we are not concerned about information spillover between the two feedback settings a subject is treated with because they do not receive any feedback under None that might affect their decisions in the Partial or Full feedback setting.

Each experimental session consisted of four decision-making parts and a questionnaire. The full details of the experimental design (including screenshots of the experimental software) are available in a supplemental appendix available from the authors.

2.3.1 Part I: TPG Game with None Feedback (20 Rounds)

Part I consisted of twenty rounds of the TPG game with None feedback. Each subject was randomly assigned into a group of three and then participated in twenty rounds with those same subjects for the entire twenty rounds. During each round, each subject made two decisions. The primary decision is to contribute or not contribute to the TPG. As is standard in experimental economics, we used neutral language to describe the decision, i.e., we used "IN" and "OUT" in place of "contributing" and "not contributing," respectively. If two of more subjects chose IN, then each group member received a "prize;" otherwise, no prize is received. We did not mention a "public good." The subject made this IN-OUT decision by selecting either the IN radio button or the OUT radio button.

The secondary decision is to report their belief about the contributions of the other two subjects in their three-subject group. Each participant assigned a percentage likelihood to each of the following possible outcomes: zero other contributors, one other contributor,

and two other contributors. Truthful reporting is incentivized via a prediction payoff with a larger payoff received for reporting beliefs that are closer to the realized contributions. The elicited-beliefs data enable us to test whether subjects are playing single-period best responses as is assumed by the theory. Their decisions are made official by clicking a "Next" button that is activated after inputting beliefs that sum to 100% and choosing either IN or OUT.

The IN-OUT and belief-reporting decisions for a single round were made on a single decision screen. Figure 2.2 shows the decision screen for a Part II round with High value ($v = 4$) and Full feedback. In Part I with None feedback, the "Outcomes" columns in the right part of the history box were entirely removed, so the history box only reported the subject's own selections.

At the conclusion of the entire experimental session, one of these twenty rounds was randomly selected for payment based on a subject's contribution decision and overall group's public good provision, and another round was randomly selected for payment based on the correctness of a subject's beliefs. We randomly choose two rounds to mitigate subjects hedging on their reported beliefs and contribution decision.

Each subject was given a fixed amount of time to complete their decisions. This timer lasted 60 seconds for rounds one through five, 50 seconds for rounds six through ten, and 40 seconds for the remaining ten rounds. If a subject did not submit their decisions before the timer expired, then the computer randomly selected IN or OUT for their contribution decision with equal likelihood (and their beliefs were set at 0% for each of the three possible outcomes). If such a round was selected for the belief payment, then a belief payment of \$0 was given. If such a round was selected for the IN-OUT decision payment, then the payment was determined by the random IN-OUT computer selection and the overall group outcome.

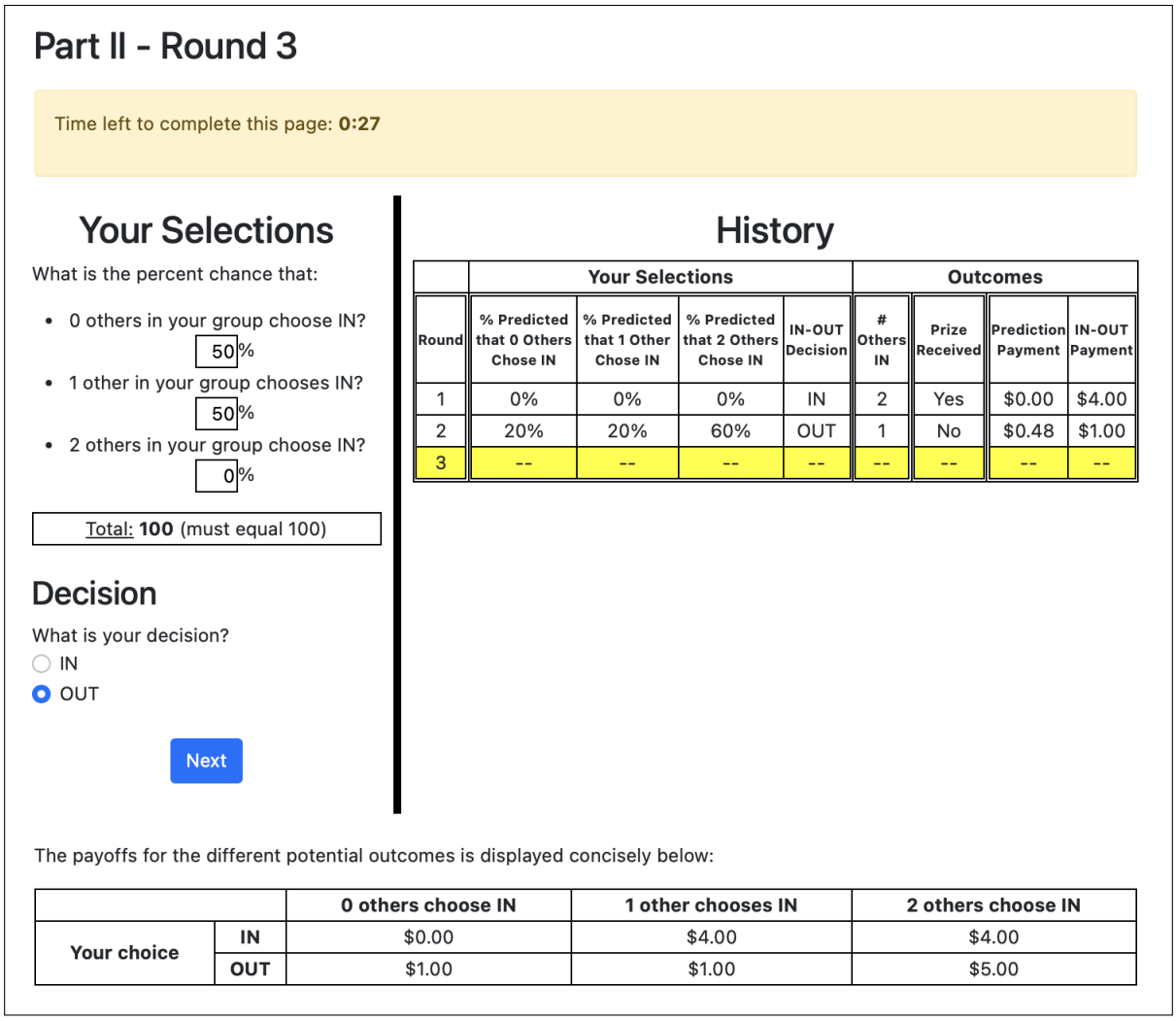


Figure 2.2: Decision Screen with High Value, $v = 4$, and Full Feedback
The Partial feedback condition is identical except the "# Others In" and "Prediction Payment" columns are entirely removed. The None feedback condition has all "Outcomes" columns removed.

2.3.2 Part II: TPG Game with Partial/Full Feedback (20 Rounds)

Part II was identical to Part I except for the change in feedback provided in the on-screen history table. Two of the twenty rounds were randomly selected from Part II (with replacement) at the end of the experimental session for the prediction and IN-OUT payments, as in Part I.

Approximately half of the subjects completed twenty rounds with Partial feedback, in which

the decision-screen history box only reported "Prize received" and "IN-OUT Payment" for the "Outcomes" columns. The other half of subjects completed twenty rounds with Full feedback which consisted of all columns shown in Figure 2.2. The twenty rounds of Part II were done without knowledge of the earnings from Part I (because those are not revealed until the end of the session).

2.3.3 Part III: The Risk-elicitation Task

Part III is a risk-elicitation task. Following a similar framework for measuring risk aversion as established in Binswanger (1980), we had the subjects choose one of five risky lotteries.

- **Option 1:** A 50% chance of earning $2.00 and a 50% chance of earning $2.00 (the sure-thing).
- **Option 2:** A 50% chance of earning $3.00 and a 50% chance of earning $1.50.
- **Option 3:** A 50% chance of earning $4.00 and a 50% chance of earning $1.00.
- **Option 4:** A 50% chance of earning $5.00 and a 50% chance of earning $0.50.
- **Option 5:** A 50% chance of earning $6.00 and a 50% chance of earning $0.00 (the highest expected payoff).

This task is done without knowing the payments from Parts I and II of the experiment to prevent payoff spillover. It is also incentivized, i.e., after the subject selected their lottery, then the computer randomly selected one outcome based on the probabilities given, and the subject is paid the respective outcome.

2.3.4 Part IV: The Charity-dictator Task

Part IV is a dictator game with a charity of the subject's choice. The subject is presented with information about four different charities: Amnesty International, the United Nations Childrens Fund (Unicef), Doctors without Borders, and the American Cancer Society. The subject selected one of these four charities and then decided how much of $3 to give to that charity, keeping the remainder for themselves.

This task is done without knowing the payments from Parts I, II, and III of the experiment to prevent payoff spillover. The amount a subject gave to the charity they selected was donated to the corresponding charity after we ran all our experimental sessions.

2.3.5 Questionnaire

After the end of Part IV, each subject answered a six-page questionnaire. The first page asked the subject to report their age, gender, race, first language, and college major. The second page showed their selections from their history box in Part I (None feedback treatment) and presented a list of strategy descriptions. They then selected at least one strategy description from the list that best matched the actions they made. A third page gave the subject a free-response text box so they can offer a description of their choices in Part I in their own words. The fourth and fifth pages were the same as the second and third pages, respectively, except they were for Part II decisions (Partial or Full feedback treatment). The sixth page asked subjects to complete three Cognitive Reflection Task questions (Frederick, 2005).

After completing the questionnaire, the subject was shown a summary results screen that reported and explained the payments received for each part of the experiment.

2.4 Hypotheses and Moderating Factors

The following hypotheses were provided in our pre-registration. Our main hypothesis predicts that strategy profiles that are Conjectural Equilibria in one experimental feedback condition but not another will be more likely to occur in the former.

Hypothesis 1. *For each value of the public good:*

(a) The three-contributor Conjectural Equilibrium will occur with higher frequency under Partial than under Full feedback.

(b) The one-contributor Conjectural Equilibrium will occur with higher frequency under None than under Partial or Full feedback.

Evidence in support of Hypothesis 1 supports an argument in favor of using Conjectural Equilibrium to generate predictions for settings with limited feedback. However, there are various moderating factors that might reduce the chances of finding strong support for Hypothesis 1, and various features of our experimental design are intended to help us better understand the reasons we find support or reject Hypothesis 1. The remaining hypotheses examine these possibilities in more detail.

Our next three hypotheses test three key underlying assumptions of Conjectural Equilibrium. One assumption is that beliefs decrease in accuracy as the feedback decreases. If this is not the case, then an equilibrium concept that explicitly considers how beliefs depend on feedback will not be very useful. Another assumption is that individuals play best responses to their beliefs no matter their feedback level. If this is not the case, then an equilibrium concept that assumes that players play best responses to their beliefs will not be useful. A third assumption is that subjects are more likely to stay in a strategy profile when that profile is a Conjectural Equilibrium than when it is not a Conjectural Equilibrium. If this is not the

case, then the Conjectural Equilibrium concept is not useful as a predictor of steady states for different levels of feedback.

Hypothesis 2. *For each value of the public good, the accuracy of subjects' beliefs decreases as feedback decreases from Full to Partial to None.*

Hypothesis 3. *For each value of the public good and feedback condition, the frequency of payoff-maximizing behavior given reported beliefs is the same.*

Hypothesis 4. *For each feedback condition and value of the public good, the frequency with which a subject changes their contribution decision from round $t-1$ to round t is lower if their group played a Conjectural Equilibrium profile for that feedback setting in round $t-1$.*

Observe that Hypotheses 3 and 4 may be rejected because of an important moderating factor, namely, the use of dynamic strategies by the subjects. In a learning environment, both the Conjectural Equilibrium and Nash Equilibrium concepts are interpreted as long-run steady states after learning and experimentation has occurred, i.e., after several periods of interaction. However, during Part I or Part II of the experiment, subjects might continue to try out different strategies to learn about others actions (i.e., experimentation) or play history-dependent strategies (e.g., conditionally cooperate) in an attempt to influence the future decisions of other subjects. Such dynamic strategies involve choosing an action that is not a single-period best response as is assumed by both Conjectural Equilibrium and Nash Equilibrium, so Conjectural Equilibrium predictions may fail because a steady state has not been reached. Our analysis will consider this possibility, and we will also ask subjects about their use of dynamic strategies.

Our final two hypotheses are concerned with other possible moderating factors. The first of these is attitudes towards risk. The Conjectural Equilibrium concept assumes that individuals' risk attitudes are constant across feedback settings, but it may be that subjects'

risk attitudes are specific to the feedback condition.[3] If so, then we may find differences in behavior across feedback conditions for a reason other than what is suggested by the Conjectural Equilibrium concept, namely, inaccurate beliefs. Our fifth hypothesis tests whether risk preferences cause behavioral changes across feedback conditions.

Hypothesis 5. *The impact of risk preferences on contribution rates does not differ by feedback condition.*

Risk attitudes may have other moderating effects on behavior. For example, a highly risk averse individual may be willing to contribute even if they believe it highly likely that there are two other contributors because their contribution reduces the risk of not providing the public good. If all group members are highly risk averse, then three-contributor profiles may be common even under Full feedback, and they might be common under Partial or None because of the risk aversion and not the limited information. To control for these possibilities, we will control for subjects' risk aversion in our regression analysis.

Our last moderating factor is the presence of social preferences. Like risk aversion, social preferences may lead to behavior that is not expected-payoff maximizing, and the social preferences may interact with the feedback treatment. For example, fairness concerns may be highly salient to subjects under Full feedback but less salient when they do not learn as detailed of information about others' contributions under Partial or None feedback. Our sixth hypothesis tests whether social preferences affect behavior differentially across feedback conditions.

Hypothesis 6. *The impact of other-regarding preferences on contribution rates does not differ by feedback condition.*

Social preferences may also affect contributions similarly across all treatments (e.g., an other-

[3]We focus here on risk aversion, but there is some theoretical work on the impact of ambiguity aversion on learning and belief formation in limited-feedback settings, e.g., Battigalli et al. (2019) and Battigalli et al. (2019).

Table 2.1: Session-by-session Breakdown

Session#	Feedback Treatment	Value Treatment	# Subjects	Average Payoff
Session 1	None / Full	Low	24	$17.95
Session 2	None / Partial	Low	21	$19.06
Session 3	None / Full	High	21	$21.11
Session 4	None / Partial	High	15	$25.39
Session 5	None / Partial	High	15	$22.04
Session 6	None / Full	Low	18	$17.55
Session 7	None / Partial	Low	18	$17.45
Session 8	None / Partial	High	15	$21.51
Session 9	None / Full	High	9	$22.29
Session 10	None / Full	High	15	$18.65

regarding subject might always contribute regardless of their beliefs), so we will control for social preferences in our regression analysis.

2.5 Results

2.5.1 Descriptive Statistics

Table 2.1 summarizes the different sessions, including the different treatments, number of subjects, and average payoffs. As expected, the average payoffs are lower in the Low public good value sessions than in the High public good value sessions. Table 2.2 provides summary statistics for the demographics and characteristics of the subjects in our experiment, e.g., the majority of our subjects are Female, Asian, and native English speakers, and many subjects are in a STEM field. As every session involved a None feedback treatment, the corresponding columns provide the aggregate summary statistics for subjects that were in the Low or High public good value treatment. Average risk aversion and charity donation are greater in the High public good value treatment than in the Low, though these differences are not statistically significant.

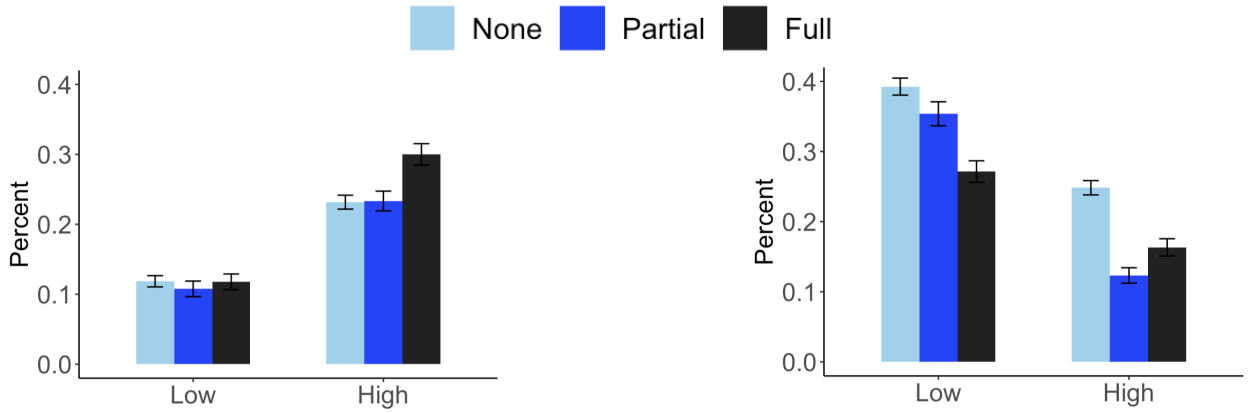
Table 2.2: Demographics and Characteristics

Feedback Treatment	None		Partial		Full	
Value Treatment	Low	High	Low	High	Low	High
Total Subjects	81	90	39	45	42	45
Demographics						
Female	77.8%	74.4%	74.4%	80.0%	81.0%	68.9%
White	19.8%	20.0%	12.8%	20.0%	26.2%	20.0%
Asian	56.8%	56.7%	53.8%	53.3%	59.5%	60.0%
Latinx Ethnicity	25.9%	26.7%	25.6%	33.3%	26.2%	20.0%
English as 1st Language	58.0%	65.6%	59.0%	62.2%	57.1%	68.9%
Avg Age	21.19	20.58	20.95	20.09	21.40	21.07
	(2.52)	(3.06)	(2.08)	(2.21)	(2.87)	(3.69)
Education						
STEM	45.7%	45.6%	41.0%	28.9%	50.0%	62.2%
Math-involved	43.2%	33.3%	33.3%	37.8%	52.4%	28.9%
Economics/Business	25.9%	22.2%	20.5%	31.1%	31.0%	13.3%
Characteristics						
Avg Risk-Aversion	2.65	2.92	2.69	3.02	2.62	2.82
	(1.48)	(1.50)	(1.45)	(1.47)	(1.51)	(1.54)
Avg Charity Donation	1.63	1.89	1.62	1.94	1.63	1.84
	(1.19)	(1.07)	(1.22)	(1.09)	(1.18)	(1.07)
Avg CRT Score	1.70	1.71	1.62	1.42	1.79	2.00
	(1.24)	(1.20)	(1.29)	(1.12)	(1.20)	(1.22)

Table 2.3: Rates of Cooperation and Good Provision

Value Treatment	Feedback Treatment	Contribution Rate	Provision Rate
Low	None	0.47	0.45
		(0.01)	(0.02)
	Partial	0.46	0.46
		(0.02)	(0.03)
	Full	0.47	0.50
		(0.02)	(0.03)
High	None	0.64	0.71
		(0.01)	(0.02)
	Partial	0.69	0.85
		(0.02)	(0.02)
	Full	0.66	0.76
		(0.02)	(0.02)

Table 2.3 shows the individual contribution rate and public good provision rate (i.e., the rate of meeting or exceeding the threshold) for each treatment. The contribution and provision rates are higher in the High public good value treatment for each feedback condition. The provision rate is also higher under Partial feedback than Full for the High public good value, but it is lower for the Low public good value.



(a) Percent of outcomes with 3 contributors

(b) Percent of outcomes with 1 contributors

Figure 2.3: Percent of Outcomes with Over and Under Contributions

2.5.2 Hypothesis 1

Evidence in support of Hypothesis 1 is mixed. Hypothesis 1(a) predicted that three contributions should occur more frequently under Partial and None than under Full, but this is rejected by a simple examination of means in Figure 2.3(a). Contrary to the prediction, three-contributor profiles occur more frequently under Full feedback than under Partial or None when the value is High, and they occur at the same rate when the value is Low. However, consistent with the prediction in Hypotheses 1(b), we see in Figure 2.3(b) that one-contributor profiles occur more often under None than Partial or Full for both Low and High public good values.

The regression analysis in Table 2.4 adds some nuance in favor of supporting Hypothesis 1. By the simple test proposed in our pre-registration, we find that three-contributor profiles occur more frequently under None relative to Full feedback in the Low public good value treatment (Regression 2), but when we interact the average risk aversion within a group with the feedback treatment (not in the pre-registration), we find that three contributors occurs more frequently for risk neutral groups under both None and Partial with a Low public good value (Regressions 5-6). For a High public good value, three-contributor profiles never occur more frequently for None and Partial than Full. Thus, by adding additional controls, we confirm Hypothesis 1(a) for Low but still reject it for High. We do note that

Table 2.4: Feedback Conditions on Over and Under-contribution

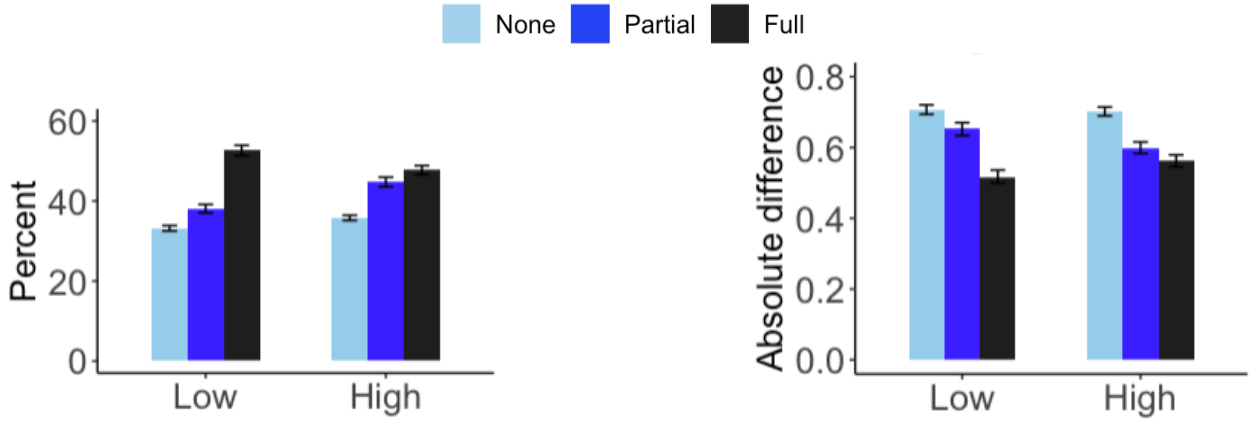
	Dependent variable:							
	(a) Three contributors							
					Interacting Group Risk Aversion			
	Low		High		Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
None	0.043	0.062**	−0.060	−0.064	0.154*	0.220**	−0.024	−0.008
	(0.028)	(0.028)	(0.043)	(0.042)	(0.088)	(0.095)	(0.116)	(0.121)
Partial	0.005	0.044	−0.053	−0.061	0.235*	0.580***	0.090	0.149
	(0.031)	(0.034)	(0.049)	(0.054)	(0.124)	(0.166)	(0.144)	(0.179)
Avg group charity donation		0.009		0.028		0.013		0.028
		(0.018)		(0.031)		(0.020)		(0.031)
Avg group CRT score		−0.020		−0.014		−0.035*		−0.009
		(0.018)		(0.029)		(0.019)		(0.030)
Avg group risk aversion		0.027*		0.053***	0.068**	0.087***	0.082**	0.079**
		(0.016)		(0.020)	(0.027)	(0.029)	(0.033)	(0.036)
None × Avg group risk aversion					−0.043	−0.059	−0.015	−0.019
					(0.037)	(0.038)	(0.042)	(0.043)
Partial × Avg group risk aversion					−0.087*	−0.199***	−0.053	−0.070
					(0.047)	(0.062)	(0.050)	(0.061)
Constant	0.064***	−0.371	0.267***	0.319	−0.115**	−0.498**	0.035	0.119
	(0.021)	(0.248)	(0.036)	(0.273)	(0.058)	(0.250)	(0.090)	(0.323)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	540	540	600	600	540	540	600	600
Adjusted R^2	0.002	0.039	0.0003	0.048	0.013	0.058	0.016	0.047
	(b) One contributor							
					Interacting Group Risk Aversion			
	Low		High		Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Partial	−0.023	−0.050	−0.170***	−0.175***	−0.509**	−0.718***	0.022	0.131
	(0.054)	(0.057)	(0.039)	(0.043)	(0.224)	(0.253)	(0.145)	(0.164)
Full	−0.200***	−0.178***	−0.147***	−0.140***	−0.261*	−0.131	−0.177	−0.269*
	(0.048)	(0.049)	(0.040)	(0.041)	(0.154)	(0.166)	(0.134)	(0.141)
Avg group charity donation		−0.023		0.025		−0.030		0.027
		(0.040)		(0.034)		(0.039)		(0.034)
Avg group CRT score		−0.059*		−0.006		−0.039		0.002
		(0.034)		(0.032)		(0.034)		(0.032)
Avg group risk aversion		0.012		−0.013	−0.023	−0.022	−0.019	−0.003
		(0.029)		(0.021)	(0.042)	(0.044)	(0.030)	(0.032)
Partial × Avg group risk aversion					0.182**	0.247***	−0.062	−0.100**
					(0.083)	(0.092)	(0.043)	(0.050)
Full × Avg group risk aversion					0.023	−0.017	0.010	0.045
					(0.055)	(0.061)	(0.044)	(0.047)
Constant	0.410***	0.135	0.290***	0.268	0.471***	0.221	0.344***	−0.007
	(0.032)	(0.339)	(0.027)	(0.270)	(0.117)	(0.367)	(0.091)	(0.320)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	496	496	565	565	496	496	565	565
Adjusted R^2	0.028	0.053	0.035	0.043	0.033	0.066	0.037	0.050

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are group averages/proportions of *age, male, white, asian, latinx, native English speakers, STEM, econ/business, and math-involved majors*. Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors are used. Results in this table use only the last ten rounds of data for each treatment to capture steady states.

average risk aversion within a group has a pronounced impact on the likelihood of a three-contributors outcome, and including the extra controls (e.g, Regression 8 vs Regression 4) shifts the coefficients on None and Partial in the direction supporting our hypothesis, despite not being significantly different from zero.

The regression results in Panel (b) of Table 2.4 provide further evidence largely in support of Hypothesis 1(b). By the tests proposed in our pre-registration (Regressions 1-4), we find that one contribution occurs less frequently under Full feedback than None regardless of the public good value. It also occurs less frequently under Partial feedback than None for a High public good value, but there is no difference in frequency across feedback settings for a Low public good value. When including the interaction of average group risk aversion with feedback treatment (Regressions 5-8), we find that one contribution occurs less frequently in Full feedback than None for a High public good value and less frequently in Partial feedback than None for a Low public good value.

Our regressions use the last ten rounds because the Conjectural Equilibrium concept is meant to capture steady states after actors have finished learning and experimenting with strategies. Our results are still mixed but slightly more supportive of Hypothesis 1 when including the last fifteen rounds or all rounds. Additionally, we drop group-level observations where at least one subject had their contribution decision made randomly by the computer due to the enforcement of a hard timer. As Hypothesis 1 measures a group-level outcome, we only include distinct observations for each group (i.e., the group outcome is not included for each subject within a group). Results are robust to using a logistic regression (see our supplemental appendix).



(a) Average percent predicted on actual outcome

(b) Average absolute difference between expected and actual outcome

Figure 2.4: Accuracy of Subjects' Beliefs

Table 2.5: Feedback Conditions on Belief Accuracy

	Dependent variable:							
	Percent on Actual Outcome				Absolute Difference in Expected and Actual Outcome			
	Low		High		Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Partial	3.650***	3.802***	8.230***	8.296***	−0.041*	−0.039*	−0.094***	−0.090***
	(1.258)	(1.270)	(1.353)	(1.386)	(0.023)	(0.023)	(0.020)	(0.021)
Full	18.268***	18.075***	10.507***	10.525***	−0.180***	−0.182***	−0.113***	−0.119***
	(1.463)	(1.453)	(1.289)	(1.302)	(0.023)	(0.023)	(0.021)	(0.021)
Risk aversion		−0.029		−2.076***		0.008		0.017***
		(0.406)		(0.373)		(0.007)		(0.006)
Charity donation		−0.439		−0.499		0.005		0.0002
		(0.456)		(0.533)		(0.008)		(0.009)
CRT questions correct		3.100***		−0.817*		−0.037***		0.014*
		(0.465)		(0.488)		(0.008)		(0.008)
Constant	34.867***	40.227***	37.380***	45.251***	0.694***	0.471***	0.673***	0.654***
	(0.678)	(5.389)	(0.642)	(4.741)	(0.013)	(0.094)	(0.012)	(0.086)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	3,145	3,145	3,503	3,503	3,145	3,145	3,503	3,503
Adjusted R^2	0.059	0.072	0.023	0.034	0.020	0.031	0.011	0.020

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are *age, female, white, asian, latinx, native English speaker, STEM, econ/business, and math-involved major*. Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. We drop observations where subjects' decision to contribute was made randomly by the computer due to the enforcement of a hard timer, as these subjects did not have beliefs recorded. Heteroskedasticity robust standard errors are used.

2.5.3 Hypothesis 2

Visual confirmation of Hypothesis 2 is seen in Figure 2.4. Figure 2.4(a) displays the average percent that subjects predicted on the outcome that actually occurred. This percent is increasing from None to Partial to Full, indicating an increase in belief accuracy as feedback increases. Figure 2.4(b) displays the absolute difference between subjects' belief of the expected number of other contributors in their group and the actual number of other contributors. This difference is decreasing from None to Partial to Full, again revealing that subjects' expected outcomes are closer to actual outcomes as feedback increases.

Table 2.5 presents regression results proposed in our pre-registration that formally test and confirm Hypothesis 2. Relative to None, the percent subjects predict on the actual outcome is greater under Partial feedback regardless of public good value, and this percent is higher still under Full feedback for both public good values (Regressions 1-4). The absolute difference between subjects' expected and actual outcome also decreases from None to Partial to Full feedback (Regressions 5-8). The coefficients on Full are also significantly different from the coefficients on Partial in most regressions indicating that beliefs are most accurate under the most feedback. Overall, Hypothesis 2 is strongly confirmed.

2.5.4 Hypothesis 3

We reject Hypothesis 3 that subjects played best responses at equal rates across the three feedback conditions. Figure 2.5 breaks downs the percent of risk-neutral best responses by the first 10 rounds, last 10 rounds, and all rounds. Subjects clearly best respond at higher rates under Full than Partial and None for both values of the public good. However, there is an increase in best responding as the rounds progress, and the last 10 rounds are driving the significant difference in the frequency of payoff-maximizing behavior across feedback conditions. Our regression results in Table 2.6 tell the same story. After controlling for

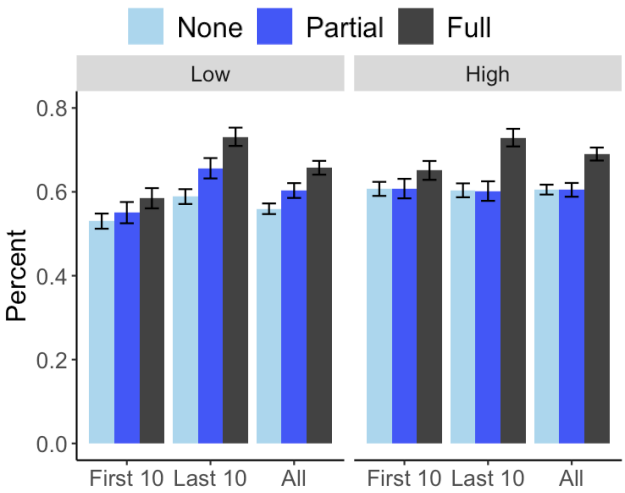


Figure 2.5: Percent of Best Responses Based on Round Number

subjects' characteristics, the rate of payoff-maximization is still statistically higher under Full feedback than None for a Low public good value. Hypothesis 3 is strongly rejected.[4]

2.5.5 Hypothesis 4

Figure 2.6 shows the frequency with which subjects changed their contribution decision when in a Conjectural Equilibrium strategy profile and when not in a Conjectural Equilibrium strategy profile in the prior round. As Hypothesis 4 predicts, subjects change their decision more frequently when not in a Conjectural Equilibrium profile. Note that every outcome in the None feedback treatment is a Conjectural Equilibrium, so this provides a baseline proclivity for how often subjects change their contribution decision. Further evidence supporting Hypothesis 4 is given in Table 2.7. Regardless of feedback condition and public good value, being in a Conjectural Equilibrium strategy profile has a strongly significant negative effect on the likelihood a subject changes their contribution decision in the following round.[5]

[4]The results in Table 2.6 are robust to logistic regressions (see our supplemental appendix). We drop observations for subjects who had their contribution decision made randomly by the computer (i.e., we dropped a group's observation if at least one member had a computer-generated random contribution decision).

[5]The results provided in Table 2.7 use all but the last round in each treatment, as we care about the overall stability of subjects' contribution decisions relative to the outcome experienced. Results are robust to a logistic regression (see our supplemental appendix). Results are robust to dropping the observations for

Table 2.6: Feedback Condition on Best Response Rate

	Dependent variable:			
	Best response			
	Low		High	
	(1)	(2)	(3)	(4)
Partial	0.068**	0.062**	−0.002	0.001
	(0.030)	(0.030)	(0.029)	(0.029)
Full	0.143***	0.146***	0.126***	0.123***
	(0.028)	(0.028)	(0.027)	(0.027)
Risk aversion		0.003		−0.028***
		(0.008)		(0.008)
Charity donation		−0.029***		−0.007
		(0.010)		(0.011)
CRT questions correct		0.046***		−0.003
		(0.011)		(0.011)
Constant	0.589***	0.726***	0.604***	0.691***
	(0.018)	(0.122)	(0.017)	(0.100)
Controls	No	Yes	No	Yes
Observations	1,575	1,575	1,762	1,762
Adjusted R^2	0.014	0.053	0.012	0.028

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are *age, female, white, asian, latinx, native English speaker, STEM, econ/business, and math-involved majors*. Results are robust to grouping Non-binary/ Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors are used. Results in this table use only the last ten rounds of data for each treatment.

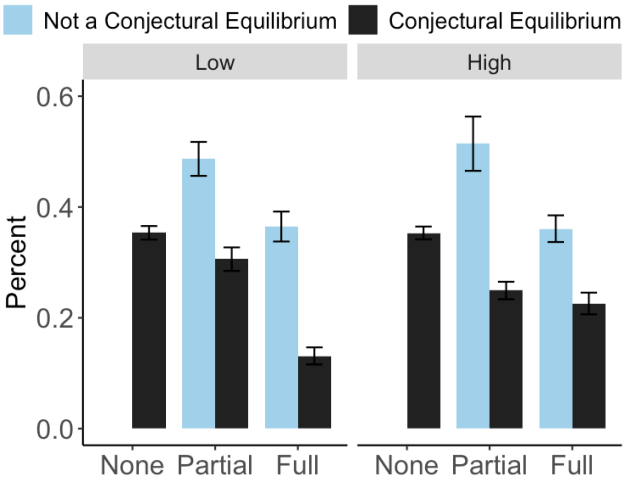


Figure 2.6: Percent of Changed Contribution Decisions

That we rejected Hypothesis 3 but confirmed Hypothesis 4 warrants closer investigation,

subjects whose decision to contribute was made randomly by the computer.

Table 2.7: Conjectural Equilibrium and Changed Contribution Decision

	Dependent variable:							
	Changed contribution decision							
	Partial				Full			
	Low		High		Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conjectural equilibrium	−0.181***	−0.192***	−0.265***	−0.231***	−0.234***	−0.200***	−0.135***	−0.143***
	(0.037)	(0.036)	(0.051)	(0.052)	(0.031)	(0.033)	(0.031)	(0.030)
Risk aversion		0.003		−0.017		−0.015		−0.002
		(0.015)		(0.011)		(0.011)		(0.012)
Charity donation		0.021		0.074***		−0.017		0.099***
		(0.017)		(0.016)		(0.014)		(0.015)
CRT questions correct		−0.002		−0.008		−0.023*		0.002
		(0.017)		(0.017)		(0.013)		(0.015)
Constant	0.487***	1.332***	0.514***	0.134	0.365***	−0.151	0.361***	0.160
	(0.031)	(0.195)	(0.049)	(0.189)	(0.027)	(0.177)	(0.024)	(0.164)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	741	741	855	855	798	798	855	855
Adjusted R^2	0.031	0.074	0.036	0.108	0.074	0.106	0.021	0.099

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are group averages/proportions of *age, male, white, asian, latinx, native English speakers, STEM, econ/business, and math-involved majors*. Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors are used. Results in this table use all but the last round of data, as subjects cannot change their contribution decision after the last round.

and one explanation is different rates of convergence to equilibrium under Partial than Full feedback. Because the Partial-feedback subjects have less information on which to update their beliefs than the Full-feedback subjects, they cannot accurately update their beliefs as quickly as the Full subjects, and this may delay their convergence to a steady state. As a result, subjects may actually be playing best responses less frequently under Partial than Full, not because their behavior is fundamentally different under Partial than Full, but instead because under Partial they are taking longer to experiment and learn before settling into a steady state.

Table 2.8 provides some evidence in support of this explanation. The negative coefficient on round number in Regression 1 indicates that subjects are changing their beliefs less as the rounds progress under the Low public good value, indicating that their beliefs are converging. When round number is interacted with feedback in Regressions 2 and 3, we see that convergence is happening most strongly under Full feedback. Consistent with that belief

convergence, for a Low public good value, groups are initially less likely to have an individual change their contribution decision under Full (Regressions 7-9), and this likelihood of at least one individual changing their contribution decreases faster under Full as the rounds progress (Regressions 8-9) even when holding fixed whether the group members played a Conjectural Equilibrium profile in the prior round. All of this evidence supports the story that convergence to a steady state is slower under Partial than Full when the public good value is Low. At face value, these findings suggest that we might eventually observe a higher rate of best responding behavior under Partial-Low if subjects were allowed to play more rounds.

However, the fact that changes in beliefs were smaller under Partial than under Full (Regressions 4-6) when the public good value is High runs counter to this story. We believe that there are two reasons for this. First, Partial-feedback subjects had higher provision rates than Full-feedback subjects (see Table 2.3). Second, many subjects appear to be maximizing the likelihood of provision rather than maximizing their expected payoff. Figure 2.7 reports the percent of subjects who selected different strategy descriptions to explain their contribution decisions (subjects could select more than one option). A large majority of subjects selected "I tried to maximize the chance that the prize will be received" (s1), which is much more than the two options that corresponded to single-period best responding (s3 and s4). Maximizing provision (rather than expected payoff) is an understandable goal under Partial because feedback only describes provision and not others' contributions. We suspect that the Partial-High subjects who focus on provision are frequently in groups with two and three-contributor profiles for which they—unlike Full-feedback subjects—cannot update their beliefs due to the Partial feedback that cannot separately distinguish between those profiles. In this instance, the beliefs stop changing sooner in Partial-High than Full-High, not because they are closer to the truth but rather because they are not able to discern how to change them. This is also consistent with what we find in Regressions 10-12, where groups are initially less likely to have an individual change their contribution decision under Full or Partial feedback than

Table 2.8: Convergence in Beliefs and Decisions across Feedback Conditions

	Dependent variable:											
	(a) Absolute change in mean beliefs						(b) At least one subject in group changed contribution decision					
	Low			High			Low			High		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Partial	0.046***	−0.024	−0.031	−0.068***	0.006	0.014	−0.157***	−0.121	−0.111	−0.189***	−0.164**	−0.126*
	(0.016)	(0.035)	(0.034)	(0.015)	(0.032)	(0.032)	(0.042)	(0.081)	(0.083)	(0.037)	(0.076)	(0.075)
Full	−0.015	0.060	0.063*	−0.017	0.002	−0.005	−0.374***	−0.289***	−0.280***	−0.247***	−0.166**	−0.180**
	(0.016)	(0.038)	(0.038)	(0.014)	(0.030)	(0.030)	(0.040)	(0.084)	(0.084)	(0.042)	(0.078)	(0.078)
Absolute difference in prior round expected and actual outcome	0.135***	0.133***	0.131***	0.101***	0.101***	0.104***						
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.015)						
Conjectural equilibrium in prior round							−0.352***	−0.346***	−0.323***	−0.253***	−0.250***	−0.247***
							(0.040)	(0.041)	(0.042)	(0.045)	(0.045)	(0.044)
Round number	−0.005***			−0.002			−0.007***			−0.007**		
	(0.001)			(0.001)			(0.003)			(0.003)		
None × Round number		−0.005***	−0.005***		0.001	0.001		−0.004	−0.004		−0.004	−0.004
		(0.002)	(0.002)		(0.002)	(0.002)		(0.004)	(0.004)		(0.004)	(0.004)
Partial × Round number		0.002	0.002		−0.006***	−0.006***		−0.007	−0.007		−0.006	−0.006
		(0.002)	(0.002)		(0.002)	(0.002)		(0.005)	(0.005)		(0.005)	(0.005)
Full × Round number		−0.012***	−0.012***		−0.001	−0.001		−0.012**	−0.012**		−0.012**	−0.012**
		(0.002)	(0.002)		(0.002)	(0.002)		(0.005)	(0.005)		(0.005)	(0.005)
Risk aversion			−0.005			−0.016***						
			(0.005)			(0.004)						
Charity donation			0.024***			0.035***						
			(0.005)			(0.005)						
CRT score			0.005			−0.014***						
			(0.005)			(0.005)						
Avg group risk aversion									0.018			−0.006
									(0.021)			(0.017)
Avg group charity donation									−0.002			0.097***
									(0.029)			(0.024)
Avg group CRT score									−0.041*			0.034
									(0.022)			(0.026)
Constant	0.235***	0.234***	0.032	0.236***	0.213***	0.250***	1.158***	1.117***	0.706***	1.054***	1.022***	0.401*
	(0.020)	(0.024)	(0.068)	(0.017)	(0.020)	(0.053)	(0.054)	(0.065)	(0.249)	(0.057)	(0.062)	(0.224)
Controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2,911	2,911	2,911	3,286	3,286	3,286	870	870	870	1,024	1,024	1,024
Adjusted R^2	0.052	0.056	0.086	0.030	0.031	0.067	0.120	0.120	0.133	0.051	0.051	0.100

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are *age, female, white, asian, latinx, native English speaker, STEM, econ/business, and math-involved majors* in Panel (a) and group averages/proportions of the same variable in Panel (b). Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors are used. We drop observations affected by a subjects' decision to contribute being made randomly by the computer. Results in Panel (b) are robust to logistic regressions.

None, but only under Full is the likelihood of at least one individual in a group changing their contribution decision decreasing faster as the rounds progress (Regressions 11-12).

2.5.6 Hypothesis 5

The regression results in Table 2.9 lead us to reject Hypothesis 5 that risk preferences do not impact contribution rates differently across treatments. To test this, we interact risk

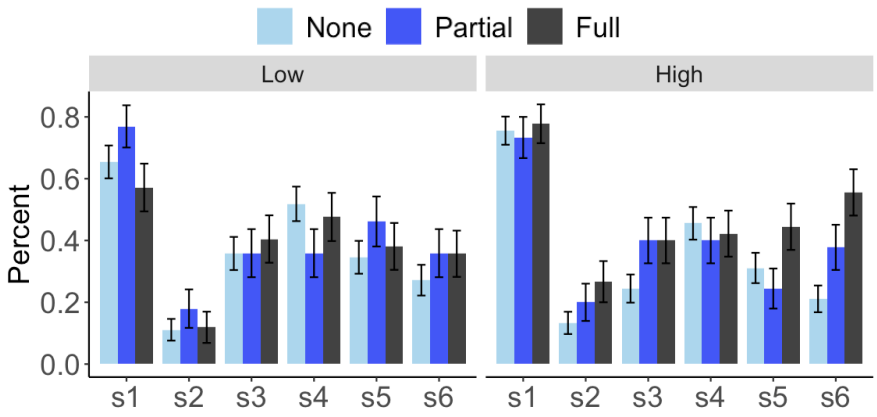


Figure 2.7: Percent of Subjects Reporting Using Strategy
Notes: s1 ="I tried to maximize the chance that the prize will be received."
s2 = "I tried to get others to pay for the prize without having to pay for it."
s3 = "I wanted to get my highest payoff in each round, not being concerned with future rounds."
s4 = "I tried to only select In if I thought there was a high chance of exactly one other subject going In."
s5 = "I tried to minimize the chance that I was the only group member who selected In."
s6 = "My decisions in some rounds were intended to affect my group members' selections in future rounds.
A seventh option "Other" is not shown because it was only selected by four subjects."

preferences on the Partial and Full feedback treatments. If Hypothesis 5 were true, we would expect these interaction terms to be statistically insignificant, but we instead find that this is the case for only Partial feedback relative to None under the Low public good value. Under the High public good value after controlling for subjects' beliefs and characteristics, greater risk aversion had a larger impact on the likelihood a subject contributes to the public good under Partial feedback relative to None. Regardless of public good value or regression controls, we find that the impact of risk aversion under Full feedback is greater relative to None.[6]

Table 2.9: Risk Aversion and Contribution Decision

	Dependent variable:					
	Contributed					
	Low			High		
	(1)	(2)	(3)	(4)	(5)	(6)
Partial	−0.060	−0.036	−0.024	0.023	−0.137**	−0.159***
	(0.064)	(0.059)	(0.058)	(0.063)	(0.059)	(0.059)
Full	−0.167***	−0.114**	−0.125***	−0.131**	−0.164***	−0.135***
	(0.060)	(0.048)	(0.048)	(0.060)	(0.053)	(0.049)
Risk aversion	0.004	0.007	−0.014	0.014	0.015	0.011
	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Partial × Risk aversion	0.007	−0.003	−0.010	0.016	0.043**	0.053***
	(0.021)	(0.019)	(0.019)	(0.019)	(0.017)	(0.017)
Full × Risk aversion	0.055***	0.030*	0.038**	0.062***	0.053***	0.043***
	(0.020)	(0.016)	(0.016)	(0.018)	(0.016)	(0.016)
Prediction on 1 other contributing		0.008***	0.008***		0.009***	0.008***
		(0.0003)	(0.0004)		(0.0004)	(0.0004)
Prediction on 2 others contributing		0.003***	0.003***		0.006***	0.005***
		(0.0004)	(0.0004)		(0.0004)	(0.0004)
Charity donation			0.052***			−0.024**
			(0.010)			(0.010)
CRT questions correct			−0.035***			0.025**
			(0.010)			(0.011)
Constant	0.451***	0.037	0.161	0.572***	0.005	0.218**
	(0.037)	(0.038)	(0.116)	(0.036)	(0.042)	(0.101)
Controls	No	No	Yes	No	No	Yes
Observations	1,575	1,575	1,575	1,762	1,762	1,762
Adjusted R^2	0.007	0.192	0.232	0.019	0.152	0.189

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are *age, female, white, asian, latinx, native English speaker, STEM, econ/business, and math-involved major*. Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors used. Results in this table use only the last ten rounds of data for each treatment.

Table 2.10: Other-regarding Preferences and Contribution Decision

	Dependent variable:					
	Contributed					
	Low			High		
	(1)	(2)	(3)	(4)	(5)	(6)
Partial	0.003	−0.025	−0.025	0.046	−0.044	−0.071
	(0.051)	(0.047)	(0.048)	(0.052)	(0.050)	(0.049)
Full	0.173***	0.023	0.034	−0.091*	−0.075	−0.047
	(0.052)	(0.046)	(0.045)	(0.055)	(0.047)	(0.043)
Charity donation	0.060***	0.054***	0.064***	−0.062***	−0.037***	−0.038***
	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)
Partial × Charity donation	−0.026	−0.011	−0.016	0.015	0.021	0.037
	(0.026)	(0.024)	(0.024)	(0.025)	(0.023)	(0.023)
Full × Charity donation	−0.119***	−0.036	−0.036*	0.071***	0.032	0.019
	(0.026)	(0.022)	(0.021)	(0.026)	(0.023)	(0.022)
Prediction on 1 other contributing		0.008***	0.008***		0.008***	0.008***
		(0.0004)	(0.0004)		(0.0004)	(0.0004)
Prediction on 2 others contributing		0.004***	0.003***		0.006***	0.006***
		(0.0004)	(0.0004)		(0.0004)	(0.0004)
Risk aversion			−0.006			0.035***
			(0.008)			(0.007)
CRT questions correct			−0.032***			0.026**
			(0.010)			(0.011)
Constant	0.361***	−0.040	0.101	0.730***	0.138***	0.163
	(0.030)	(0.031)	(0.114)	(0.031)	(0.042)	(0.100)
Controls	No	No	Yes	No	No	Yes
Observations	1,575	1,575	1,575	1,762	1,762	1,762
Adjusted R^2	0.015	0.200	0.230	0.014	0.135	0.185

Notes: *p<0.1; **p<0.05; ***p<0.01. Controls are *age, female, white, asian, latinx, native English speaker, STEM, econ/business, and math-involved major*. Results are robust to grouping Non-binary/Prefer not to answer with female rather than with male observations. Heteroskedasticity robust standard errors are used. Results in this table use only the last ten rounds of data for each treatment.

2.5.7 Hypothesis 6

Hypothesis 6 predicts that other-regarding preferences should not impact how often a subject contributes differently across the feedback conditions. Using donations to charity as a proxy

[6]The results in Table 2.9 use the last ten rounds of each treatment to test the equilibrium prediction of Hypothesis 5. We also drop observations where subjects' decision to contribute was made randomly by the computer due to the enforcement of a hard timer. Results are robust to a logistic regression (see our supplemental appendix).

for other-regarding preferences, the regression results in Table 2.10 provide moderate support for Hypothesis 6. If Hypothesis 6 is true, we expect to see the interaction between charity donation and feedback treatment to be statistically insignificant. We find this to be the case under the High public good value treatment after controlling for subjects' beliefs. We further find that the impact of other-regarding preferences on contribution rates is not different under the Low public good value treatment for Partial feedback relative to None. The only exception is that under the Low public good value there is evidence at the ten percent significance level that higher other-regarding preferences decreases the likelihood a subject contributes under Full feedback relative to None.[7]

2.6 Discussion

Our goal was to test the predictive power of Conjectural Equilibrium in a limited-feedback setting. To this end, we have found a variety of support in favor of using the Conjectural Equilibrium concept but with some notable evidence that also recommends some caution. Here, we summarize and comment on our key findings.

Foremost, we find evidence that supports the two main motivating justifications for Conjectural Equilibrium. First, limited feedback hinders individuals' ability to form accurate beliefs as evidenced by reported beliefs becoming less accurate as feedback decreases (Hypothesis 2 confirmed). Second, these inaccurate beliefs can lead individuals to persist in choosing actions that are Conjectural Equilibria even if they are not Nash Equilibria (Hypothesis 4 confirmed). These two findings provide immediate justification for choosing the Conjectural Equilibrium concept instead of the Nash Equilibrium concept when making predictions for limited-feedback settings. The Conjectural Equilibrium allows for less accurate beliefs in

[7]The regression results in Table 2.10 use the last ten rounds of each treatment to test the equilibrium prediction of Hypothesis 6. We again drop observations where subjects' decision was made randomly by the computer. Results are robust to a logistic regression (see our supplemental appendix).

equilibrium when there is limited feedback, which in turn can lead to steady states in which individuals are playing non-Nash Equilibrium strategies.

However, there are three moderating factors that complicate the use of Conjectural Equilibrium for predictions. The first is that outcomes are very sensitive to risk aversion, which implies that making precise predictions using the Conjectural Equilibrium Concept requires a knowledge of the subjects' risk preferences (see Hypothesis 1 discussion). More problematically, risk aversion interacts positively with feedback (Hypothesis 5 rejected) such that more risk averse subjects are more willing to contribute as feedback goes from None to Full. This interactive effect is sufficiently strong that we had only mixed success in predicting outcomes under limited feedback (see Hypothesis 1 discussion).

The second moderating factor is a slower convergence to steady state under limited feedback (see Hypothesis 3-4 discussion). Our results for the Low public good value are consistent with individuals being less likely to play single-period best responses in early rounds as they experiment in their contribution decisions and learn about their group members' contributions, but then more frequently playing single-period best responses in later rounds after their beliefs about others' behavior approaches a steady state. If this assessment is correct and it takes longer to reach a steady state under limited feedback, then Conjectural Equilibrium predictions for limited-feedback settings might not be valid until after many rounds of learning, perhaps more than we had our subjects play.

The third factor is that individuals might have as their objective a goal that is not maximizing their expected payoffs. In our TPG experiment, more subjects claim to be maximizing the likelihood of provision than claim to be maximizing their payoffs (see Hypothesis 4 discussion), and if true then we would expect a high rate of three-contributor outcomes even when it is not a Conjectural Equilibrium (or Nash Equilibrium) because three-contributor outcomes have provision. As a result, we would not see an increase in three-contributor outcomes when switching from Full to Partial because the rate of three-contributor outcomes

was already high. The Conjectural Equilibrium concept that assumes expected-payoff maximization will be less useful when subjects are not attempting to maximize their payoffs. Of course, this shortcoming of Conjectural Equilibrium applies equally to Nash Equilibrium.

These findings reveal both promise and limitations when applying the Conjectural Equilibrium concept. Theorists are correct when they assert that the persistence of incorrect beliefs is an important equilibrium phenomenon when individuals have limited feedback, but their theory should account for how feedback interacts with individual preferences over risk. Experimentalists should use Conjectural Equilibrium rather than Nash Equilbrium to obtain predictions in limited-feedback settings, but they should allow subjects more time to converge to a steady state when there is limited feedback for sharper predictions.

Our findings also suggest some fruitful directions for future research. Our experimental study is limited to a TPG game, so future work should test whether our findings from the TPG game are consistent with what is found in other games. While we study a TPG game where subjects only decide to contribute or not, allowing subjects to decide how much of an endowment to contribute might offer further insight into the impacts of risk aversion and limited feedback. Additional work is also needed to measure how much extra time is needed for convergence to steady state when individuals have limited feedback. While we test for subjects' risk and other-regarding preferences, incorporating a measure for ambiguity aversion may also yield useful results. Future studies along these lines will provide additional clarity on the value of Conjectural Equilibrium as a predictive equilibrium concept.

Chapter 3

Cybersecurity measures and incident frequency: Evidence from the UK

3.1 Introduction

With most of the world today having an online presence, cybersecurity is an increasingly relevant concern for governments, businesses, and individuals. Total losses from cybercrime reached record highs in 2023 at $12.5 billion, with an average of over 50,000 data breaches and over 300,000 phishing incidents reported each year since 2021 (FBI, 2023). As a result, investment in security and risk management has continued to sharply rise (FBI, 2023; Moore, 2022) in an effort to combat cybersecurity attacks. While cybersecurity investment is important for protecting against attacks, there is not much understanding about what cybersecurity measures are most effective in reducing the likelihood and frequency of cybersecurity incidents. Understanding which measures are the most beneficial in reducing incident frequencies is important for all organizations that choose to invest in cybersecurity.

In this paper, I assess the impact of various cybersecurity measures on the frequency of

cybersecurity incident.[1] Employing a new panel dataset from the Cybersecurity Longitudinal Survey (CSLS), I am able to analyze these relationships using a fixed effects regression approach. The CSLS is a survey of randomly sampled organizations across the United Kingdom that collects information on cybersecurity measures adopted and incidents experienced. As this paper is the first exploration of the CSLS, the analysis focuses on developing an understanding of the relationships between cybersecurity measures and incident frequency.

In presenting descriptive statistics of incident frequency, I provide new firm-level stylized facts regarding the persistence of cybersecurity incidents. Notably, I show that there is some degree of persistence in an organization's experienced incident frequency. While this persistence is still largely apparent for restricted samples that consider non-phishing incidents or only phishing incidents, a higher proportion of organizations transition to experiencing no incidents after previously experiencing some number of incidents. This is suggestive that organizations may be adopting cybersecurity measures that both deter potential attackers and reduce the number of serious incidents.

Specifically regarding the effectiveness of cybersecurity measures, I find that rules for moving and storing data, restricting work-related personal device usage, and staff cybersecurity training are all associated with large decreases in incident frequency, especially for relatively more serious incidents (i.e., non-phishing). In practice, this means an organization may experience more favorable cybersecurity outcomes by password-protecting sensitive data, providing devices to employees for offsite work, and requiring annual cybersecurity training (or phishing detection exercises), respectively.[2] To my knowledge, the relevance of data rules and personal devices are novel empirical findings, though they are consistent with theoretical predictions in Deurlington (2024), which presents a model exploring organizational connectivity

[1] Namely, I consider how measures are related to experiencing incidents in the following categorical frequencies: ever, more than once, and monthly or more. Additionally, to understand how these relationships change depending on the relative severity of incidents, I consider outcomes for the frequency of the following incident categories: all types, all types excluding phishing, and only phishing.

[2] Note that these are a few examples, and there are many more possible avenues an organization could take instead.

and attacker and defender behavior. Further, my results corroborate the findings in Buil-Gil et al. (2022), Celeny et al. (2024), and McCrohan et al. (2010) that training staff in cybersecurity is important.

I additionally find that monitoring user activity and formally assessing the cybersecurity risks posed by suppliers and partners is associated with reductions in the likelihood and frequency of phishing incidents. These are novel findings on how cybersecurity measures are related to phishing incident frequency. These results begin paving the way for understanding what deters attackers from targeting an organization in the first place.[3]

The results I find are especially beneficial for organizations and policymakers interested in mitigating pertinent cybersecurity threats for a few reasons. First, I identify cybersecurity measures that are associated with lower incident frequencies. Also, I outline possible avenues for deterring potential attackers in the first place. Given my novel findings on the efficacy of implementing data storage rules, restrictions on work-related personal device usage, monitoring user activity, and auditing business partners' cybersecurity, organizations and policymakers should seek to find methods of promoting the implementation of these measures.[4]

This paper primarily contributes to two literatures. The first of these is the literature estimating the effectiveness of cybersecurity measures (Gandal et al., 2023; Buil-Gil et al., 2022; Dambra et al., 2020; Aldasoro et al., 2022). These papers explore how implementing specific measures reduce the likelihood of experiencing a cybersecurity incident. My contribution to this literature is three-fold. First, my paper considers three different thresholds of incident frequency when estimating the effectiveness of cybersecurity measures, rather than only considering whether or not an organization experienced incidents. It is often unrealistic for organizations to expect no incidents altogether, so there is value to understanding what

[3]Related to Deurlington (2024), these results are suggestive that monitoring access points to digital assets raises the relative cost of attacking and may deter some attackers from targeting organizations with these defensive measures.

[4]While the importance of staff cybersecurity training is not a novel result, it is also an effective measure that should be promoted by organizations and policymakers.

measures reduce the overall frequency with which incidents occur. Second, I consider three groups of incident classification to assess how cybersecurity measures are related to the likelihood of experiencing different severities of incidents rather than only considering all incident classifications together. Measures may affect the likelihood of non-phishing and phishing incidents differently, yet this difference has not been explored in this literature. Finally, I will be using a panel dataset, allowing me to capture time and firm-level fixed effects. To my knowledge, mine is the first paper to use organization-level panel data to determine the relationship between cybersecurity measures and incident frequency.[5]

The second literature I contribute to is regarding cybersecurity in the United Kingdom. There have been a handful of papers exploring a government-run cross-sectional survey known as the Cybersecurity Breaches Survey (CSBS)[6] (De Arroyabe et al., 2023; Kemp et al., 2023; Heitzenrater and Simpson, 2016; Buil-Gil et al., 2022). These papers study an array of topics, which include the effectiveness of cybersecurity measures as well as what factors into organizations' decisions to report incidents or invest in cybersecurity. My paper is the first to utilize panel data from the CSLS. Panel data for a random sample of organizations across a nation's economy is rare in cybersecurity surveys, so the CSLS offers a unique opportunity to explore cybersecurity in the UK.

My paper also generally contributes to the quickly growing field of economics and cybersecurity. Literature in this field aims to understand and measure costs in the context of cybersecurity (Anderson et al., 2013, 2019; Moore, 2010; Agrafiotis et al., 2018) as well as explain specific aspects of cyber risk and how organizations are affected by an evolving threat space (Edwards et al., 2016; Chidukwani et al., 2022; Alawida et al., 2022). Despite an understanding of the

[5]Hawdon et al. (2020) considers rates of cyber victimization using survey panel data for individuals, so their analysis is not relevant at the firm-level. Other papers have analyzed time series data of cybercrime, but do not assess the effectiveness of cybersecurity measures in reducing cybercrime (Buil-Gil et al., 2021; Kemp et al., 2021). In exploring whether organizations should invest more in cybersecurity, Dinkova et al. (2023) points out the need for panel data to identify the effectiveness of cybersecurity measures.

[6]The CSBS collects cross-sectional data on cybersecurity decisions and experiences for randomly sampled organizations across the UK.

prevalence of phishing and its effectiveness as an initial vector for cyberattacks (Chiew et al., 2018; FBI, 2023), none of these papers have added to our understanding of how organizations are affected by phishing incidents and what they can do to mitigate risks posed by phishing. Using data from the CSLS, I am able to fill in this gap in the literature.

The rest of this paper is structured as follows. Section 2 introduces the CSLS data and provides summary statistics. Section 3 outlines the empirical strategy. Section 4 presents and discusses results. Section 5 concludes and discusses paths for future research.

3.2 Data Description

In this paper, I analyze data from the first two waves (2021-2022) of the Cybersecurity Longitudinal Survey (CSLS).[7] The CSLS is a survey of medium and large businesses and high-income charities in the United Kingdom.[8] This survey is commissioned by the Department for Science, Innovation and Technology (DSIT) and the Department for Digital, Culture, Media and Sport (DCMS) and developed in conjunction with Ipsos. Survey administrators collect information on a wide range of organizations' cybersecurity measures and experiences with cyber incidents.

The CSLS offers many strengths compared to other surveys on cybersecurity in organizations. A random probability sampling approach is used to avoid selection bias, and Random Iterative Method (RIM) weighting is used to account for non-response bias and skewed sampling on business size and sector. The CSLS includes both online and telephone data collection to include organizations with a smaller online presence and reduce self-selection bias. While most cybersecurity surveys are cross-sectional, the longitudinal nature of the

[7]Data from the third and final wave of the CSLS has yet to be made publicly accessible.

[8]Researchers interested in cybersecurity in small businesses, low-income charities, and educational institutions should refer to the CSBS (Department for Science, Innovation and Technology and Department for Digital, Culture, Media and Sport, 2017).

CSLS allows researchers to study how organizations' cybersecurity posture and incident experiences change over time. Additionally, the CSLS distinguishes between phishing and non-phishing incidents, allowing researchers to consider outcomes of varying degrees of severity.

While the survey collects data on an array of cybersecurity aspects, there is minimal data on general firm-level characteristics (i.e., revenue) and other relevant data (i.e., IT expenditures). Furthermore, as with any survey, there are limitations in the data that should be mentioned. Because organizations are self-reporting statistics, there will likely be under-reporting in both the scope and severity of any incidents experienced for several reasons. Organizations can only report cyber incidents they have identified, may not track costs associated with incidents, or may wish to look more favorable in their responses (even though answers are assured to remain confidential and anonymous). Therefore, results in this paper should be considered lower-bounds for frequency and severity.

In 2021, the CSLS surveyed 1205 businesses and 536 charities. Of these, 435 businesses and 239 charities were surveyed again in 2022 (along with a new set of 253 businesses and 134 charities added to refresh the sample). The remainder of this paper will focus only on these 674 organizations that were surveyed in both years. Table 3.1 provides additional firm-level descriptive statistics. As this paper is focused on the organizations surveyed in both 2021 and 2022, much is the same between these years. However, there are some changes in firm sizes, and three organizations in the "Education" or "Service or membership" sectors in 2021 became registered charities in 2022. To see more details regarding the survey methodology and sampling, the reader is referred to Department for Digital, Culture, Media and Sport (2022) and Department for Digital, Culture, Media and Sport (2023).

Table 3.2 presents a summary of cybersecurity measures taken by organizations in 2021 and 2022.[9] The first column identifies a measure within one of five categories: General Protection,

[9]To conserve space and improve readability, Table 3.2 presents descriptive statistics for only the most

Table 3.1: Organization-level Characteristics

	2021		2022	
Classification				
Business	438	65.0%	435	64.5%
Charity	236	35.0%	239	35.5%
Size				
Under 50	0	0.0%	11	1.6%
50 to 249	305	45.3%	299	44.4%
250 to 499	63	9.3%	56	8.3%
500 to 999	43	6.4%	41	6.1%
1,000 or more	27	4.0%	27	4.0%
Unknown	0	0.0%	1	0.1%
Charity (no size given)	236	35.0%	239	35.5%
Sector				
Administration	49	11.2%	49	11.3%
Arts or recreation	11	2.5%	11	2.5%
Construction	20	4.6%	20	4.6%
Education	18	4.1%	16	3.7%
Finance or insurance	19	4.3%	19	4.4%
Food or hospitality	36	8.2%	36	8.3%
Health, social care or social work	35	8.0%	35	8.0%
Information or communication	35	8.0%	35	8.0%
Manurfacturing	86	19.6%	86	19.8%
Profession, scientific or technical	34	7.8%	34	7.8%
Real estate	3	0.7%	3	0.7%
Retail or wholesale	60	13.7%	60	13.8%
Service or membership	6	1.4%	5	1.1%
Transport or storage	22	5.0%	22	5.1%
Utilities or production	4	0.9%	4	0.9%
Charity	236	53.9%	239	54.9%

Rules and Policies, Incident Management, Vulnerability Identification, and Visibility.[10] The next four columns illustrate how organizations changed their cybersecurity measures from 2021 to 2022 (i.e., 184 organizations did not have staff training in 2021 and in 2022). The third of these four columns shows that a non-negligible number of organizations stop implementing a cybersecurity measure in 2022 that they had in 2021. Though it is not immediately intuitive why an organization would drop a cybersecurity measure, there could be several reasons for doing so. Namely, the organization's management may have found

empirically relevant cybersecurity measures. For a more complete descriptive statistics table of cybersecurity measures taken by organizations, see Table C.1 in Appendix C.

[10]These groupings are similar to those discussed in Buil-Gil et al. (2022), which analyzes the cross-sectional CSBS dataset.

Table 3.2: Cybersecurity Measures Summary

Has measure in 2021	No	No	Yes	Yes	Other		
Has measure in 2022	No	Yes	No	Yes	Other	2021	2022
General Protection							
Training in past 12 months	184	108	50	314	18	54.5%	63.2%
AI or ML tools	314	51	40	82	187	20.2%	22.7%
Rules and Policies							
Any monitoring of user activity	119	77	81	349	48	65.6%	66.3%
Rules for storing and moving files containing personal data	22	45	57	513	37	87.2%	85.0%
All five Cyber Essentials[1]	157	112	101	304	0	60.1%	61.7%
Incident Management							
Business Continuity Plan	72	53	42	436	71	74.3%	75.5%
Risk register	121	61	61	323	108	61.3%	61.3%
Written list of IT estate and vulnerabilities	103	82	83	290	116	59.8%	59.5%
Incident Response Plan	164	90	54	298	68	54.2%	60.2%
Vulnerability Identification							
Formally assessed risks presented by any partners	326	70	56	106	116	26.0%	29.5%
Visibility							
Staff can access network/files through personal devices	259	84	114	207	10	48.2%	43.3%
Staff can connect to network/files outside workplace	127	47	62	388	50	68.5%	67.1%
Has a VPN for staff connecting remotely	119	45	72	420	18	73.9%	69.7%
Uses a cloud server that stores data/files	93	84	69	410	18	71.8%	74.2%
Uses a physical server that stores data/files	104	27	52	479	12	79.5%	75.5%

[1]Firewalls, secure configurations, access controls, malware protection, and patch management

the measure too expensive to implement (with respect to finances or time) or believed that the measure was simply not useful. The sixth column captures the number of organizations that did not provide a definitive "No" or "Yes" in both 2021 and 2022 to having adopted a measure. For example, a firm that reports "No" for training in 2021 and "Don't know" in 2022. Analysis of these cybersecurity measures later in the paper will largely ignore organizations that do not report a "No" or a "Yes" in both 2021 and 2022. Finally, the last two columns provide an aggregate summary of the proportion of firms in the respective year that adopted a cybersecurity measure. There generally seems to be higher rates of adoption in 2022 with the exception of Visibility. Some definitions of terms found in Table 3.2 and in later tables can be found in Table C.4 in Appendix C.

Details on the types of incidents organizations faced is presented in Table 3.3. Relatively few organization experienced the same type of incident in both 2021 and 2022, aside from phishing and being impersonated. The last two columns present the proportion of organizations

that reported the corresponding incident type in 2021 and 2022, respectively. Note that the same organization could report multiple types of incidents, and each organization that reports an incident type is only counted once even if they experienced this type of incident many times.[11] Especially noteworthy is that over 70 percent of organizations experienced at least one phishing incident in each year. No other types of incidents were experienced by a majority of organizations. Furthermore, aside from incidents involving an organization being impersonated, no other types of incidents were experienced by more than 12 percent of organizations. As a result of the commonality of phishing, my analysis will distinguish what cybersecurity measures are associated with specifically phishing incident frequency in addition to overall incident frequency.

Table 3.3: Types of Incidents Experienced

Occurred in 2021	No	No	Yes	Yes		
Occurred in 2022	No	Yes	No	Yes	2021	2022
Ransomware	607	25	15	2	2.5%	4.2%
Malware/viruses	523	51	42	23	10.1%	11.4%
Unauthorised internal use (staff)	582	26	17	10	4.2%	5.9%
Unauthorised external use	588	24	29	2	4.7%	4.2%
Denial of service attacks	568	24	25	12	5.8%	5.6%
Attempted hacking - online bank accounts	595	15	10	5	2.2%	3.6%
Attempted hacking - website, social media, or accounts	506	42	25	33	11.7%	11.4%
Impersonating organization	274	91	84	184	40.8%	41.8%
Phishing	89	88	63	411	71.1%	75.4%
Unauthorised listening into video conferences	622	2	6	1	1.0%	0.4%
Other	550	34	31	11	7.0%	7.0%

Table 3.4 provides transition matrices for how frequently organizations experienced cybersecurity incidents. In other words, the row is the frequency of incident experienced in 2021 and the column is the frequency experienced in 2022. Panel (a) presents the transition matrix of reported incident frequency for all types of incidents, Panel (b) presents the transition matrix when excluding phishing incidents, and Panel (c) presents the transition matrix for only phishing incidents. As the frequency in Panel (c) is not directly reported in the data,

[11]It is also worthwhile to point out that two incidents may typically be reported in conjunction. For example, it is possible that each ransomware incident in this sample started from a phishing incident. This is not explored, nor does it take away from the results, in this paper.

Table 3.4: Changes to Incident Frequency

(a) **Including Phishing**

	None	Once	More than once	Monthly	Weekly	Daily or more	Total
None	48%	12%	21%	11%	5%	4%	164
Once	13%	11%	32%	19%	11%	13%	62
More than once	14%	12%	38%	18%	11%	7%	141
Monthly	10%	9%	25%	35%	11%	10%	145
Weekly	12%	4%	22%	21%	24%	18%	78
Daily or more	5%	8%	10%	8%	19%	50%	62
Total	133	65	167	127	78	82	652

(b) **Excluding Phishing**

	None	Once	More than once	Monthly	Weekly	Daily or more	Total
None	69%	7%	10%	6%	5%	3%	312
Once	34%	10%	24%	16%	6	10%	50
More than once	38%	7%	26%	14%	9%	6%	103
Monthly	22%	6%	21%	31%	9%	10%	88
Weekly	20%	2%	18%	22%	18%	22%	54
Daily or more	9%	9%	9%	9%	17%	49%	65
Total	307	47	103	88	54	65	664

(c) **Only Phishing**

	None	Once	More than once	Monthly	Weekly	Daily or more	Total
None	64%	5%	17%	9%	3%	2%	150
Once	38%	0%	38%	12%	12%	0%	8
More than once	33%	5%	29%	14%	14%	5%	21
Monthly	29%	11%	25%	25%	7%	4%	28
Weekly	41%	0%	18%	12%	18%	12%	17
Daily or more	20%	0%	40%	0%	20%	20%	5
Total	122	12	46	27	14	8	229

I create this using the reported incident frequency (found in Panel (a)) for the subset of organizations that reported no phishing incidents or only phishing incidents in Table 3.3. This excludes organizations that experienced both ransomware and phishing incidents, for example, as it would be unclear how frequently the organization experienced each type of incident based on the reported value given in Panel (a). Doing this allows me to infer that the frequency reported is for specifically phishing incidents if this was the only incident an

organization experienced. Note that this does substantially reduce the sample size in Panel (c) relative to Panels (a) and (b).

As can be seen in Panel (a), the modal rate for each column is on, or near, the diagonal of the transition matrix. This shows there is some degree of persistence in an organization's experienced incident frequency. However, many organizations in the sample experience a change in their incident frequency, showing this persistence is imperfect. While any change in incident frequency is possible – as each entry in the transition matrix has nonzero probability – most organizations experience within a one step change of their original frequency. Considering Panel (b), this pattern of persistence generally holds, though a higher proportion of organizations transition to no incidents in 2022 after experiencing some number of incidents in 2021. As Panel (b) excludes phishing incidents, this suggests either that (1) some organizations do a better job of preventing non-phishing incidents or (2) that cybercriminals are less interested in breaching the same organization across years. Panel (c) present similar patterns as those in Panel (b), but with more noise due to the smaller sample size. Many organizations in the sample transition to no phishing incidents in 2022, which could be explained by (1) these organizations adopting better cybersecurity measures that deters potential attackers or (2) cybercriminals taking more interest in new targets. However, the lack of large values in the upper triangle of the Panel (c) matrix does not lend much support to the idea that cybercriminals prefer new targets, as relatively few organizations experience increased rates of phishing.[12]

3.3 Empirical Strategy

The goal of this paper is to identify what cybersecurity measures reduce the likelihood and overall frequency of cyber incidents experienced by an organization. My preferred fixed

[12]To be clear, I am not dismissing this explanation, as I am using a relatively small random sample and attackers could simply be targeting new organizations outside of this sample.

effects regression specification is:

$$incident_{i,t} = \beta posture_{i,t} + \alpha_i + \gamma_t + \epsilon_{i,t}$$

where $posture_{i,t}$ is a general term for dummy variables that capture whether an organization implements various cybersecurity measures within five overarching categories: general protection, rules and policies, incident management, vulnerability identification, and visibility. α_i and γ_t respectively represent firm and year fixed effects. The outcome of interest, $incident_{i,t}$, is a binary outcome variable capturing whether or not an organization experienced cyber incidents above a specified frequency threshold. Due to limited data, especially at higher frequencies of incidents, my analysis focuses on the following three thresholds: no incidents versus at least one incident (0 vs 1-6), at most one incident versus more than one incident (0-1 vs 2-6), and incidents less than monthly versus at least monthly (0-2 vs 3-6). These thresholds can be visualized in Table 3.5.

Table 3.5: Tiered Incident Thresholds

	0 vs 1-6 (1)	0-1 vs 2-6 (2)	0-2 vs 3-6 (3)
Several times a day (6)	$incident_i = 1$		
Roughly once a day (5)			
Roughly once a week (4)			
Roughly once a month (3)			
More than once (2)			
Once only (1)		$incident_i = 0$	
No attack (0)			

The reader should note that the first threshold (0 vs 1-6) is similar to existing literature that estimates the impact of cybersecurity measures on the likelihood of experiencing any breaches, whereas my primary contribution comes in providing the second and third thresholds to identify whether certain cybersecurity measures reduce the frequency of incidents irrespective of their effect on the preventing incidents entirely. That is, if the coefficient on some measure is insignificant for the first threshold but significant for thresholds two or three, the measure

is effective in reducing the overall frequency of incidents an organization experiences but is not likely to help in preventing all incidents.

To establish the importance of including firm-level fixed effects and justify my model specification, I present an example in Table 3.6 regressing only staff training on the incident frequency outcomes. Not accounting for firm fixed effects in Regressions 1-9 shows that staff cybersecurity training is associated with an increased likelihood of high incident frequencies. However, while size and sector fixed effects may generally capture what types of organizations are valuable to cybercriminals, they do not provide any critical information in terms of how likely an organization is to be targeted that would not be captured with firm-level fixed effects. For example, a large (size) manufacturing (sector) firm may be more or less valuable to attackers depending on the specific product it manufactures (i.e., semiconductors). This detail is addressed when including firm fixed effects, but not when only accounting for size and sector fixed effects. Regressions 10-12 take include firm fixed effects, and the results are consistent with my preferred model specification, which includes firm and year fixed effects.[13] After accounting for firm fixed effects, we observe that staff cybersecurity training is instead associated with a decreased likelihood of high frequencies of incidents.

Table 3.6: Differing Fixed Effects Regressions

	Dependent variable: Any incidents											
	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Staff training	0.096***	0.103***	0.077**	0.092**	0.098***	0.078***	0.071*	0.079***	0.052**	-0.026	-0.123**	-0.122**
	(0.019)	(0.025)	(0.031)	(0.024)	(0.010)	(0.012)	(0.034)	(0.016)	(0.016)	(0.047)	(0.036)	(0.041)
Size FEs	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FEs	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Observations	1,147	1,147	1,147	1,147	1,147	1,147	1,147	1,147	1,147	1,147	1,147	1,147
Adjusted R^2	0.020	0.013	0.005	0.028	0.023	0.016	0.064	0.046	0.046	0.439	0.224	0.213

*p<0.1; **p<0.05; ***p<0.01. All results are robust to including year fixed effects. Regressions 10-12 are robust to dropping size and sector fixed effects.

Furthermore, to assess the impact of a cybersecurity measure on reducing the frequency of more or less serious incidents, I assess three related outcome variables. First, the reported frequency of all types of incidents, which includes phishing. Second, the reported frequency

[13]Year fixed effects should be included in the model specification to account for annual changes in cybersecurity investments and threats (FBI, 2023; Moore, 2022).

of incidents when organizations exclude phishing. And third, the reported frequency of only phishing incidents. As discussed earlier, the frequency of phishing incidents was derived using an organization's reported frequency of incident combined with their reported types of incidents experienced. For threshold (1) in Table 3.5, the outcome is simply a one if an organization reported experiencing phishing incidents in the past 12 months and a zero otherwise. This corresponds to an organization reporting phishing in Table 3.3. However, for thresholds (2) and (3), the outcome is a one if an organization reported experiencing incidents more frequently than the corresponding threshold and did not report incidents other than phishing. This means I will be using the smaller sample shown in Panel (c) of Table 3.4 for thresholds (2) and (3).

There are two primary assumptions to consider with this empirical approach. First is the assumption of strict exogeneity. I consider the effect of cybersecurity measures within broad categories – general protection, rules and policies, incident managment, vulnerability identification, and visibility – so as to reduce time-variant omitted variable bias (as including firm fixed effects addresses time-invariant omitted variable bias).[14] However, organizations report both cybersecurity measures and experiences at the same time in the CSLS, meaning they may have adopted new measures after experiencing one or more incidents earlier in the year. As a result, causal interpretation of a cybersecurity measure's effect on incident frequency will be limited and instead, analysis will focus on the association of these variables. The second assumption is regarding rank deficiency. As seen in Table 3.2, there is a moderate degree of heterogeneity and change across most cybersecurity measures. Therefore, the main concerns of rank deficiency will come from too small of sample sizes. I attempt to mitigate this by only including small groups of cybersecurity measures in the regressions together to limit the number of observations dropped due to null responses.

Despite the potential limitations arising from this empirical strategy – specifically those

[14]There is minimal correlation between cybersecurity measures that are not in the same category.

related to the timing and nature of survey data collection – the fixed effects regression approach offers a distinct advantage. Namely, controlling for firm fixed effects, I am able to control for time-invariant omitted variables that other papers in this literature have been unable to do as a result of their using cross-sectional data. An especially important feature of this strategy is the ability to capture organizations' inherent value to cybercriminals. An organization's inherent value to attackers is arguably time-invariant, as organizations tend to have a fixed nature and mission, and changes to their digital presence tend to be slow. For example, a bank will continue to be valuable to cybercriminals year-to-year due to the value of the financial information they hold.

3.4 Results

In this section, I present results expressing the relationship between organizations' cybersecurity measures and incident frequency. Specifically, each subsection corresponds to the groupings of measures presented in areas of general protection, rules and policies, incident management, vulnerability identification, and visibility. Note that each cybersecurity measure examined is a binary variable that captures whether or not an organization responded to the CSLS that they have the corresponding measure. Additionally, results explore the impact of implementing cybersecurity measures on incident frequency for varying degrees of incident severity. The primary contributions of the following results are twofold: (1) understanding how cybersecurity measures are related to incident frequency in addition to incident likelihood, and (2) understanding how cybersecurity measures are related to phishing incidents in addition to all types of incidents.

3.4.1 General Protection

Table 3.7 provides results considering the relationship between general cybersecurity measures taken and incident frequency. These measures include cybersecurity training, the use of cybersecurity tools that use AI or ML, compliance with standards and accreditations, and cyber insurance protection. None of these measures are strongly correlated with or clearly fit in the other groupings of cybersecurity measures (rules and policies, incident management, vulnerability identification, and visibility).

Table 3.7: General Protection Measures and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Staff training	0.008	-0.144**	-0.218***	0.061	-0.068	-0.177**	0.000	-0.103	-0.148
	(0.045)	(0.070)	(0.072)	(0.070)	(0.079)	(0.082)	(0.061)	(0.135)	(0.115)
AI or ML tools	0.012	-0.068	-0.035	-0.018	-0.083	-0.134	0.017	0.173	0.356**
	(0.034)	(0.069)	(0.079)	(0.062)	(0.090)	(0.083)	(0.057)	(0.177)	(0.137)
Adheres to at least one standard/accreditation	-0.015	-0.039	-0.067	-0.053	-0.088	-0.047	-0.038	-0.013	-0.042
	(0.054)	(0.073)	(0.083)	(0.083)	(0.092)	(0.105)	(0.063)	(0.127)	(0.107)
Cyber Insurance (rel. to "specific CS insurance policy")									
CS covered as part of general policy	0.021	0.019	-0.057	0.030	0.025	-0.072	0.020	0.171	0.115
	(0.038)	(0.055)	(0.074)	(0.056)	(0.064)	(0.086)	(0.050)	(0.169)	(0.119)
Not insured against CS incidents	0.013	0.034	-0.024	-0.070	-0.139	-0.101	-0.007	0.181	0.056
	(0.063)	(0.115)	(0.111)	(0.100)	(0.131)	(0.116)	(0.067)	(0.144)	(0.104)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	779	779	779	696	696	696	883	289	289
Adjusted R^2	0.463	0.226	0.273	0.544	0.348	0.262	0.348	0.250	0.247

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

Carrying out training sessions for staff not involved in cybersecurity is associated with lower frequencies of incidents, both when including and excluding phishing (regressions 2, 3, and 6). However, organizations that trained non-IT staff still have roughly the same likelihood of experiencing no incidents relative to those that did no training. Regressions 8 and 9 also provide some suggestive evidence that employee training may be associated with lower rates of phishing incidents, though these effects are insignificant and the respective sample sizes

are relatively small. This provides evidence that organizations requiring staff training in cybersecurity experience lower incident frequencies than organizations with no cybersecurity training. Furthermore, these reductions in incident frequency are seen when assessing all incident types and when excluding phishing incidents, which suggests that staff cybersecurity training is important in reducing the incident frequency of relatively more serious incidents.

Having artificial intelligence (AI) or machine learning (ML) cybersecurity tools does not seem to matter for the frequency of non-phishing incidents experienced. With a larger sample, a researcher might detect a significant result as the direction of this relationship is primarily negative across Regressions 1-6. This implies that AI or ML tools may be related to an organization experiencing lower frequencies of more serious (i.e., non-phishing) cyber incidents. However, the results in regressions 7-9 are suggestive (and strongly significant in 9) that having AI or ML tools is associated with a greater frequency of phishing incidents. This is likely because AI or ML tools are useful for detecting phishing or because they are more likely to label certain incidents as phishing, both of which would make an organization more likely to report higher phishing frequencies.

The effect of complying to certain cybersecurity standards or obtaining accreditations is insignificant for any type or tier of incident frequency. However, the results in Table 3.7 could be suggestive of a larger pattern for compliance to cybersecurity standards, as the direction of this effect is negative across each regression. In other words, with more data, it may be possible to show that adhering to cybersecurity standards is associated with lower frequencies of incidents. There are no significant effects or easily discernible patterns, however, for the relationship between cyber insurance and incident frequency. This could suggest there is no systemic moral hazard problem arising in the UK cyber insurance market, as organizations with more specific cyber insurance policies do not experience higher rates of incidents than firms without cyber insurance.

3.4.2 Rules and Policies

The relationship between cybersecurity rules and policies with incident frequency is shown in Table 3.8. Controls for whether organizations have policies related to malware protection, firewalls, restricted access rights, or security controls on devices are not included, as the vast majority of organizations have these in place. However, the effectiveness of these measures is captured to some extent in whether or not an organization has all five Cyber Essentials (firewalls, secure configurations, access controls, malware protection, and patch management).

Table 3.8: Cybersecurity Policies and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6 (1)	0-1 vs 2-6 (2)	0-2 vs 3-6 (3)	0 vs 1-6 (4)	0-1 vs 2-6 (5)	0-2 vs 3-6 (6)	0 vs 1-6 (7)	0-1 vs 2-6 (8)	0-2 vs 3-6 (9)
Apply software updates within 14 days	0.035	-0.095	-0.147	0.070	-0.044	-0.102	-0.076	0.016	-0.076
	(0.039)	(0.080)	(0.095)	(0.095)	(0.112)	(0.116)	(0.064)	(0.109)	(0.116)
Monitoring of user activity	0.006	-0.021	-0.002	0.002	0.018	0.008	-0.054	-0.153*	-0.028
	(0.030)	(0.047)	(0.064)	(0.053)	(0.067)	(0.070)	(0.048)	(0.086)	(0.092)
Rules for storing or moving data	-0.018	-0.024	-0.144**	-0.029	-0.048	-0.130*	-0.014	-0.003	-0.091
	(0.045)	(0.068)	(0.066)	(0.060)	(0.079)	(0.072)	(0.057)	(0.110)	(0.115)
Backing up data securely via cloud service	-0.017	0.005	-0.012	0.070	0.072	0.029	-0.009	-0.096	-0.039
	(0.046)	(0.071)	(0.079)	(0.070)	(0.082)	(0.077)	(0.058)	(0.102)	(0.095)
Backing up data securely via NOT cloud service	-0.012	-0.081	-0.072	-0.028	-0.026	-0.059	-0.052	-0.118	-0.008
	(0.043)	(0.066)	(0.070)	(0.067)	(0.075)	(0.076)	(0.058)	(0.108)	(0.086)
All five Cyber Essentials	-0.038	0.109	0.160*	-0.114	0.077	0.183*	-0.003	-0.046	0.080
	(0.037)	(0.077)	(0.085)	(0.096)	(0.110)	(0.108)	(0.063)	(0.098)	(0.112)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1002	1002	1002	890	890	890	1128	359	359
Adjusted R^2	0.490	0.222	0.226	0.539	0.353	0.260	0.351	0.274	0.163

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

There is not a significant relationship between monitoring user activity and the likelihood of experiencing cyber incidents when including or excluding phishing. However, user monitoring is weakly associated with a lower likelihood of experiencing more than one phishing incident (Regression 8). Additionally, the direction of the coefficients on user monitoring in Regressions

7 and 9 provide further suggestive evidence that this measure may be associated with lower frequencies of phishing incidents. This would be a promising result supporting the idea of deterrence, as it suggests attackers avoid organizations that monitor their users so as to avoid detection. More research should be done to explore this effect.

Regressions 3 and 6 provide moderate evidence that having specific rules for moving and storing data is associated with a roughly 14 percent decrease in the likelihood of experiencing incidents monthly or more. While this effect is insignificant across the other regressions in Table 3.8, the direction of this relationship is maintained. Furthermore, the magnitude of the effect in Regression 9 being larger than in Regressions 7 and 8 is consistent with the pattern we observe in Regressions 1-3 and 4-6, indicating that data storage rules may be important in reducing high rates of incidents, but not in preventing incidents altogether.

There is a weakly significant and large positive relationship between an organization having all five of the Cyber Essentials – firewalls, secure configurations, access controls, malware protection, and patch management – and an increased likelihood of experiencing all and non-phishing incidents monthly or more (Regressions 3 and 6). While not significant, however, the direction of this relationship is reversed when the threshold for the outcome variable is any incidents (Regressions 1 and 4). Together, this might suggest the presence of two types of organizations that incorporate all five Cyber Essentials. The first of these are organizations frequently targeted by cybercriminals due to offering high potential value (rich personal or financial data, a government affiliation, etc.). These organizations adopt the five Cyber Essentials, as they are likely to be more aware of their cybersecurity risks due to the value of their digital assets, but these measures will likely be insufficient to deter attackers. The second class of organizations are those that are conscious of cybersecurity, but offer lower potential value to cybercriminals. When the value of these digital assets are low, it is possible that the five Cyber Essentials may deter some attackers and prevent incidents.

3.4.3 Incident Management

Table 3.9 gives the relationship between an organization's incident management procedures and its experienced incident level. Note that incident management has to do with how an organization is prepared to respond in the case of a breach, and it is not immediately obvious that these measures should have an effect on the likelihood of an incident. However, it is possible they are helpful in deterring continuous or repeated incidents and in preventing incidents from snowballing into larger breaches.

Table 3.9: Incident Management Measures and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6 (1)	0-1 vs 2-6 (2)	0-2 vs 3-6 (3)	0 vs 1-6 (4)	0-1 vs 2-6 (5)	0-2 vs 3-6 (6)	0 vs 1-6 (7)	0-1 vs 2-6 (8)	0-2 vs 3-6 (9)
Business continuity plan	0.053	0.020	-0.010	0.025	0.019	-0.045	0.116*	0.093	0.261***
	(0.048)	(0.087)	(0.086)	(0.066)	(0.087)	(0.087)	(0.061)	(0.125)	(0.097)
Risk register covering CS	0.044	-0.041	-0.002	0.044	-0.042	0.024	-0.039	-0.125	-0.185**
	(0.039)	(0.071)	(0.081)	(0.061)	(0.090)	(0.086)	(0.048)	(0.104)	(0.084)
Defines acceptable risk level	-0.006	0.044	0.007	-0.042	0.008	0.034	-0.027	0.088	0.022
	(0.026)	(0.049)	(0.064)	(0.050)	(0.066)	(0.070)	(0.046)	(0.108)	(0.122)
Identifies most critical assets	-0.047	-0.007	0.020	0.009	0.040	0.057	-0.055	0.050	0.033
	(0.034)	(0.050)	(0.066)	(0.058)	(0.073)	(0.076)	(0.050)	(0.105)	(0.088)
Written list of IT estate and vulnerabilities	-0.093**	-0.136***	0.008	-0.055	-0.067	0.041	-0.039	-0.126	-0.053
	(0.038)	(0.050)	(0.066)	(0.050)	(0.063)	(0.076)	(0.049)	(0.111)	(0.107)
Incident response plan	-0.037	0.085	-0.014	-0.144**	-0.052	-0.043	-0.035	0.154	0.053
	(0.032)	(0.060)	(0.068)	(0.057)	(0.079)	(0.082)	(0.053)	(0.129)	(0.116)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	910	910	910	805	805	805	1014	315	315
Adjusted R^2	0.532	0.233	0.265	0.536	0.319	0.290	0.345	0.263	0.145

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

Regressions 7-9 in Table 3.9 provide moderate evidence that having a business continuity plan is associated with a higher likelihood of experiencing more phishing incidents. The fact that this relationship is insignificant for all and non-phishing incidents (Regressions 1-6) could suggest that organizations with business continuity plans tend to offer more apparent value to cybercriminals, which is why they tend to experience more phishing. Because these

organizations experience higher rates of phishing incidents, they may be more likely to plan for worst-case scenarios (i.e., create a business continuity plan).

Though no large pattern emerges, an organization maintaining a risk register covering cybersecurity is associated with a nearly 20% decrease in the likelihood of monthly or more phishing incidents (Regressions 9). This could be due to the organization being relatively more aware of the specific risks they face, leading to them filtering out more phishing (i.e., through better spam filters) or raising awareness in staff more effectively. However, due to the small sample size in creating a variable for phishing incident frequency, minimal weight should be placed on this result. The fact that the direction of this relationship remains negative in Regressions 7 and 8, though, suggests more research could be done in parsing out the usefulness of a risk register in reducing phishing incidents.

Organizations with a written list of their IT estate and vulnerabilities appear less likely to experience more that one cyber incident (Regressions 1 and 2). Though insignificant, the direction of this relationship seems to be consistent for both non-phishing and phishing incidents. Two plausible reasons could explain this. One, it is easier for an organization to have a written list of their IT estate and digital vulnerabilities if their estate is smaller and they have fewer vulnerabilities. Having less IT and vulnerabilities may suggest an organization has less to protect from cybercriminals, meaning they get attacked less frequently and therefore suffer fewer incidents. Or two, organizations with a written list of their IT estate and vulnerabilities are better organized and understand how to monitor and protect their network more effectively. A dataset that includes the size of an organization's IT department (staff size and/or expenditures) and number of computers across the entire organization, for example, would help in assessing these explanations.

Having an Incident Response Plan (IRP) is associated with nearly a 15 percent decrease in the likelihood of experiencing any non-phishing incidents (Regression 4). Though insignificant, the direction of this effect is still negative in Regressions 5 and 6, suggesting an IRP could

be helpful in reducing the frequency of more serious (i.e., non-phishing) incidents. This relationship might be explained by accepting that IRPs indicate an organization that is more responsive and prepared for threats. If this is true, organizations with IRPs may detect and stop issues sooner, preventing more incidents. This explanation fits with the fact that having an IRP is not – or potentially even positively – related to the frequency of experienced phishing incidents (Regressions 8-9). Implementing an IRP appears to be among the most effective measures for organizations looking to improve their cybersecurity posture. In addition to potentially reducing the likelihood of non-phishing incidents, an IRP is likely to reduce the overall time and resources when dealing with incidents that do occur because a response plan has been predefined.

3.4.4 Vulnerability Identification

Table 3.10 assesses the impacts of measures an organization takes with respect to identifying vulnerabilities in its network on incident likelihood and frequency. The expected effect of identification measures is not immediately clear. On the one hand, better identification means an organization is likely to detect more threats and breaches, resulting in more recorded incidents. On the other hand, cybercriminals generally wish to go unnoticed – especially when they are first entering and exploring an organization's network – and as a result, may prefer attacking organizations with worse identification. In addition to this, adopting more vulnerability identification measures could mean an organization is better at getting ahead of potential issues and quickly patches holes in their network, leading to lower incidents in general.

No significant results or overarching patterns emerge for the relationships between the majority of vulnerability identification approaches and incident frequency. This could be due to the presence of competing effects as outlined above. The fact that we do not clearly

Table 3.10: Vulnerability Identification Measures and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Vulnerability audit	0.008	0.015	-0.025	0.019	-0.015	-0.030	0.018	0.121	-0.027
	(0.033)	(0.056)	(0.056)	(0.052)	(0.060)	(0.061)	(0.045)	(0.120)	(0.106)
Risk assessment	0.006	0.012	0.093	-0.041	-0.010	0.082	-0.011	-0.033	0.063
	(0.039)	(0.053)	(0.068)	(0.060)	(0.070)	(0.081)	(0.048)	(0.089)	(0.112)
Threat intelligence	-0.023	-0.026	-0.010	0.065	0.046	0.047	0.002	0.026	0.017
	(0.029)	(0.049)	(0.060)	(0.044)	(0.061)	(0.064)	(0.044)	(0.123)	(0.133)
Security monitoring tools	0.012	-0.032	0.035	-0.038	-0.085	-0.050	0.025	-0.052	0.110
	(0.040)	(0.065)	(0.070)	(0.061)	(0.080)	(0.078)	(0.055)	(0.102)	(0.104)
Formally assessed/managed partners' CS risks	-0.008	-0.110**	-0.003	0.038	-0.047	0.023	-0.102**	-0.141	-0.079
	(0.033)	(0.055)	(0.067)	(0.058)	(0.074)	(0.077)	(0.050)	(0.134)	(0.118)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	916	916	916	815	815	815	1023	322	322
Adjusted R^2	0.517	0.214	0.280	0.537	0.340	0.301	0.375	0.285	0.106

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

see positive relationships between the outcome variable and these identification measures suggests the possibility that identification could be useful in deterring attacks. Much more granular data on attacker and defender behaviors and incentives would be needed, however, to establish this deterrence effect.

However, there does appear to be a negative relationship between incident frequency and whether or not an organization audits or manages suppliers' and partners' cybersecurity (Regressions 2 and 7). Specifically, it seems that auditing partners tends to be associated with a reduction in phishing incidents, as the direction and magnitude of the relationships in Regression 7-9 are more noteworthy than in Regression 4-6. This could be related to the high proportion of organizations that report being impersonated in Table 3.3. Organizations that audit the cybersecurity of their suppliers and partners are more likely to identify organizations in their network that have been infiltrated or are being impersonated. As a result, they may be able to detect and prevent phishing attempts from seemingly trustworth sources in addition to helping their partners achieve secure networks again.

3.4.5 Visibility

Visibility refers to measures that could limit or expand the number of vulnerabilities within an organization's network. Rather than technical cybersecurity policies or tools, the measures that I assess in this section highlight organizational practices that influence the number of vulnerabilities present within an organization that could be exploitable. The relationship between the visibility of an organization and incident frequency is shown in Table 3.11. Data in the CSLS on organizational visibility (i.e., websites and social media) is relatively limited, though a few important variables can be assessed. Namely, an organization's personal device restrictions, choice of server for data and file storage, and requirements for remote working (if applicable).

Table 3.11: Visibility Measures and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Staff can access network/files with personal device	0.004	-0.050	0.061	0.077*	0.041	0.132**	0.017	-0.003	-0.018
	(0.023)	(0.045)	(0.058)	(0.041)	(0.056)	(0.064)	(0.030)	(0.081)	(0.074)
Uses a cloud server that stores data/files	-0.018	-0.006	0.004	0.098*	0.156**	0.049	0.026	-0.033	0.084
	(0.035)	(0.053)	(0.068)	(0.053)	(0.065)	(0.077)	(0.039)	(0.070)	(0.067)
Has a physical server that stores data/files	-0.059	-0.019	-0.083	-0.185***	0.016	-0.047	-0.092	-0.012	-0.080
	(0.049)	(0.076)	(0.084)	(0.054)	(0.086)	(0.102)	(0.056)	(0.114)	(0.076)
VPN and Remote Work (rel. to "Staff can connect without a VPN")									
No remote work	0.005	-0.035	-0.042	0.007	-0.066	0.003	-0.051	0.003	-0.114
	(0.052)	(0.070)	(0.094)	(0.078)	(0.085)	(0.104)	(0.058)	(0.111)	(0.094)
Staff forced to connect with a VPN	-0.027	-0.164***	-0.030	-0.051	-0.198***	-0.046	-0.002	0.004	-0.084
	(0.028)	(0.052)	(0.074)	(0.056)	(0.075)	(0.081)	(0.040)	(0.107)	(0.104)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1119	1119	1119	986	986	986	1273	427	427
Adjusted R^2	0.537	0.243	0.254	0.557	0.349	0.267	0.417	0.265	0.202

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

Allowing staff to access an organization's network and files through personal devices is associated with a 7.7% increase in the likelihood of any non-phishing incidents (Regression 4) and a 13.2% increase in the likelihood of monthly or more non-phishing incidents (Regression

6). Furthermore, the direction of this relationship in Regression 5 is suggestive that not allowing employees to access sensitive materials from personal devices is important for reducing the frequency of more serious incidents.

An organization's choice of server for data and file storage appears to have an important relationship with the frequency of non-phishing incidents experienced. Notably, using a cloud server is associated with a 15.6% increase in the likelihood of experiencing more than one non-phishing incident (Regression 5), while using a physical server is associated with an 18.5% decrease in the likelihood of experiencing any non-phishing incidents (Regression 4). The direction of coefficients on using a physical server are mostly negative, suggesting physical server use is related to lower incident frequencies. It is worth considering that smaller organizations and those with less need for data storage due to limited data collection (and hence, less potential value to attackers) may be both more likely to use a local physical server and less valuable targets to cybercriminals.

Interestingly, the likelihood and frequency of any type of incident does not seem to be statistically different for organizations with no remote work relative to organizations with remote work where staff are allowed to connect to the organization's network without using a virtual private network (VPN). However, organizations that require remote workers to connect with a VPN have lower likelihoods of experiencing more than one incident relative to organizations where remote workers can connect without a VPN (Regressions 2 and 5). Notably, the direction of the effect of requiring VPN connections in Regressions 1-6 is suggestive that requiring a VPN is associated with lower rates of non-phishing incidents relative to allowing staff to connect without a VPN. These results could indicate that non-security-conscious individuals reside in organizations regardless of whether or not the organization has remote work, but forcing all individuals to use a VPN while remote working might limit the risks posed by at least some of the non-security-conscious employees. Alternatively, these results could be indicative that organizations that force remote workers to connect to

the company network with a VPN are generally more security-conscious, and therefore better at preventing frequent incidents.

As an extension, I explore the relationship between the involvement of organizations' board (of directors) in cybersecurity and incident frequency. Descriptive statistics, empirical results, and a corresponding discussion are provided in Appendix C.

3.5 Discussion and Conclusion

This paper had three overarching objectives. The first of these was to assess the relationship between cybersecurity measures and incident frequency. The second objective was to distinguish these relationships for phishing incidents from all types of incidents. Finally, this paper showcased a panel dataset not yet analyzed by researchers despite the lack of available cybersecurity panel data. Here I summarize my findings, highlight limitations of my results, and discuss promising directions for future research.

I add to findings in other papers that training is important in reducing incident frequency. I also present new results showing that rules for data storage and restrictions on work-related personal device usage are associated with decreases in incident frequency. With respect to only phishing, I find that assessing supply chain risks and monitoring user activity are associated with lower likelihoods and frequencies of incidents. These are the first results showing a negative relationship between cybersecurity measures and phishing incident frequency. This provides a possible path for future research interested in estimating the deterrence effect of cybersecurity.

While I control for firm and year fixed effects in my analysis, the fact that organizations report both cybersecurity measures and experiences at the same time in the survey data lead to endogeneity concerns. Organizations may have adopted new measures after having

experienced one or more incidents early in the year, yet this timeline is not captured within the survey data. Generally, this would likely imply that not having a measure may be related to higher frequencies of incidents, meaning the results I find are lower bounds on the effectiveness of measures in reducing incident frequency. However, causal interpretation of the results in this paper should be limited without introducing an effective instrumental variable.[15]

Due to a relatively small sample size of only 674 organizations in the panel, there are generally high standard errors associated with many results. Several measures show consistent directional patterns, however, and their effectiveness in reducing incident frequency should be further explored in larger panel datasets. The measures that I believe should be further explored in larger panels are: complying with cybersecurity standards or accreditations, applying timely software updates, backing up data via cloud versus non-cloud services, writing a list of IT estate and vulnerabilities, and using a physical versus cloud server for data storage. The third wave of CSLS data will be published later in 2024, at which point the analysis in this paper should be extended and the effectiveness of the above measures should be revisited.

There are several possibilities for future research stemming from this paper and the CSLS. One, the CSLS collects data on what, and how many, resources an organization uses for cybersecurity guidance and information. This could possibly be an effective instrument for understanding the causal implications of cybersecurity measures. Two, other empirical strategies may provide useful results for measuring the effectiveness of cybersecurity. For example, a first differences approach could explore the impact of introducing or dropping a measure on cybersecurity outcomes.[16] Additionally, using principal component analysis may be interesting for determining the bundles of cybersecurity measures that are most effective when implemented together, or what types of firms opt for certain bundles of

[15]What, or how many, resources an organization uses for guidance and information on cybersecurity is collected in the CSLS. This could possibly be converted into an effective instrument.

[16]The concern of correlated errors across time periods would need to be resolved in using this approach.

measures. Finally, this paper focused on incident frequency as the outcome variable, though the CSLS also provides comprehensive data on a number of costs and damages experienced by organizations. This could be leveraged to understand the effectiveness of incident management procedures in reducing costs or to measure the costs of certain types of incidents, for example.[17]

[17]Anderson et al. (2013) suggests that there is under-investment in incident management and over-investment in attack prevention. Therefore, it would be valuable to know what measures, especially related to incident management, reduce the expected damages and impacts of breaches. Pursuing this research path could provide even stronger support in favor of organizations adopting an IRP.

Bibliography

Acemoglu, D., A. Malekian, and A. Ozdaglar (2016). Network security and contagion. *Journal of Economic Theory 166*, 536–585.

Agrafiotis, I., J. R. Nurse, M. Goldsmith, S. Creese, and D. Upton (2018). A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity 4*(1), tyy006.

Alawida, M., A. E. Omolara, O. I. Abiodun, and M. Al-Rajab (2022). A deeper look into cybersecurity issues in the wake of covid-19: A survey. *Journal of King Saud University-Computer and Information Sciences 34*(10), 8176–8206.

Aldasoro, I., L. Gambacorta, P. Giudici, and T. Leach (2022). The drivers of cyber risk. *Journal of Financial Stability 60*, 100989.

Anderson, R., C. Barton, R. Böhme, R. Clayton, C. Ganán, T. Grasso, M. Levi, T. Moore, and M. Vasek (2019). Measuring the changing cost of cybercrime. In *The 18th Annual Workshop on the Economics of Information Security (WEIS 2019)*.

Anderson, R., C. Barton, R. Böhme, R. Clayton, M. J. Van Eeten, M. Levi, T. Moore, and S. Savage (2013). Measuring the cost of cybercrime. *The economics of information security and privacy*, 265–300.

Arce, D. G., D. Kovenock, and B. Roberson (2012). Weakest-link attacker-defender games with multiple attack technologies. *Naval Research Logistics (NRL) 59*(6), 457–469.

Au, W. (2004). Criticality and environmental uncertainty in step-level public goods. *Group Dynamics: Theory, Research, and Practice 8*, 40–61.

Azrieli, Y. (2009). On pure conjectural equilibria with non-manipulable information. *International Journal of Game Theory 38*, 209–219.

Bagnoli, M. and B. Lipman (1992). Private provision of public goods can be efficient. *Public Choice 74*, 59–78.

Battigalli, P., E. Catonini, G. Lanzani, and M. Marinacci (2019). Ambiguity attitudes and self-confirming equilibrium in sequential games. *Games and Economic Behavior 115*, 1–29.

Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2015). Self-confirming equilibrium and model uncertainty. *American Economic Review 105*, 646–677.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019). Learning an self-confirming long-run biases. *Journal of Economic Theory 183*, 740–785.

Battigalli, P., M. Gilli, and M. C. Molinari (1992). Learning and convergence to equilibrium in repeated strategic interactions: An introductory survey. *Ricerche Economiche 46*, 335–378.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American journal of agricultural economics 62*(3), 395–407.

Bloch, F., B. Dutta, and M. Dziubiński (2020). A game of hide and seek in networks. *Journal of Economic Theory 190*, 105119.

Borel, E. (1921). La théorie du jeu et les équations intégralesa noyau symétrique. *Comptes rendus de l'Académie des Sciences 173*(1304-1308), 58.

Buil-Gil, D., E. Barrett, et al. (2022). The dynamics of business, cybersecurity and cyber-victimization: Foregrounding the internal guardian in prevention. In *The New Technology of Financial Crime*, pp. 5–34. Routledge.

Buil-Gil, D., F. Miró-Llinares, A. Moneva, S. Kemp, and N. Díaz-Castaño (2021). Cybercrime and shifts in opportunities during covid-19: a preliminary analysis in the uk. *European Societies 23*(sup1), S47–S59.

Celeny, D., L. Maréchal, E. Rousselot, A. Mermoud, and M. Humbert (2024). Prioritizing investments in cybersecurity: Empirical evidence from an event study on the determinants of cyberattack costs. *The 23rd Annual Workshop on the Economics of Information Security (WEIS 2024). arXiv preprint arXiv:2402.04773*.

Chen, D. L., M. Schonger, and C. Wickens (2016). otree - an open-source platform for laboratory, online and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Chidukwani, A., S. Zander, and P. Koutsakis (2022). A survey on the cyber security of small-to-medium businesses: challenges, research focus and recommendations. *IEEE Access 10*, 85701–85719.

Chiew, K. L., K. S. C. Yong, and C. L. Tan (2018). A survey of phishing attacks: Their types, vectors and technical approaches. *Expert Systems with Applications 106*, 1–20.

Chowdhury, S. M., D. Kovenock, D. Rojo Arjona, and N. Wilcox (2016). Focality and asymmetry in multi-battle contests. *Chapman University, Economic Science Institute, Working Paper*, 16–16.

Chowdhury, S. M., D. Kovenock, and R. M. Sheremeta (2013). An experimental investigation of colonel blotto games. *Economic Theory 52*, 833–861.

Clark, D. J. and K. A. Konrad (2007). Asymmetric conflict: Weakest link against best shot. *Journal of Conflict Resolution 51*(3), 457–469.

Dambra, S., L. Bilge, and D. Balzarotti (2020). Sok: Cyber insurance–technical challenges and a system security roadmap. In *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 1367–1383. IEEE.

De Arroyabe, I. F., C. F. Arranz, M. F. Arroyabe, and J. C. F. de Arroyabe (2023). Cybersecurity capabilities and cyber-attacks as drivers of investment in cybersecurity systems: A uk survey for 2018 and 2019. *Computers & Security 124*, 102954.

Deck, C. and R. M. Sheremeta (2012). Fight or flight? defending against sequential attacks in the game of siege. *Journal of Conflict Resolution 56*(6), 1069–1088.

Dekel, E., D. Fudenberg, and D. Levine (1999). Payoff information and self-confirming equilibrium. *Journal of Economic Theory 89*, 165–185.

Department for Digital, Culture, Media and Sport (2022). Cyber security longitudinal survey wave 1 technical annex. URL: `https://assets.publishing.service.gov.uk/media/61ee9fa9e90e0703796dd024/Technical\_report_-_cyber_security_longitudinal_survey.pdf`.

Department for Digital, Culture, Media and Sport (2023). Cyber security longitudinal survey wave 2 technical annex. URL: `https://assets.publishing.service.gov.uk/media/63933d00e90e0769b576c05c/Technical\_Report_-_Cyber_Security_Longitudinal_Survey_-_wave_two_accessible.pdf`.

Department for Science, Innovation and Technology and Department for Digital, Culture, Media and Sport (2017). Cyber security breaches survey. URL: `https://www.gov.uk/government/collections/cyber-security-breaches-survey`.

Deurlington, C. (2024). Defense and connectivity of weakest-link networks. *Working Paper*.

Dinkova, M., R. El-Dardiry, and B. Overvest (2023). Should firms invest more in cybersecurity? *Small Business Economics*, 1–30.

Dziubiński, M. and S. Goyal (2013). Network design and defence. *Games and Economic Behavior 79*, 30–43.

Dziubiński, M. and S. Goyal (2017). How do you defend a network? *Theoretical Economics 12*(1), 331–376.

Edwards, B., S. Hofmeyr, and S. Forrest (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity 2*(1), 3–14.

Esponda, I. (2013). Rationalizable conjectural equilibrium: A framework for robust predictions. *Theoretical Economics 8*, 467–501.

FBI (2023). Federal bureau of investigation internet crime report 2023. URL: `https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf`.

Financial Crimes Enforcement Network (2021). Ransomware trends in bank secrecy act data between january 2021 and june 2021. *FinCEN advisory, Oct*, 2021–10.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*, 25–42.

Gandal, N., T. Moore, M. Riordan, and N. Barnir (2023). Empirically evaluating the effect of security precautions on cyber incidents. *Computers & Security 133*, 103380.

Gilli, M. (1999). On non-nash equilibria. *Games and Economic Behavior 27*, 184–203.

Goyal, S. and A. Vigier (2014). Attack, defence, and contagion in networks. *The Review of Economic Studies 81*(4), 1518–1542.

Hausken, K. (2006). Income, interdependence, and substitution effects affecting incentives for security investment. *Journal of Accounting and Public Policy 25*(6), 629–665.

Hausken, K. (2017). Defense and attack for interdependent systems. *European Journal of Operational Research 256*(2), 582–591.

Hawdon, J., K. Parti, and T. E. Dearden (2020). Cybercrime in america amid covid-19: The initial results from a natural experiment. *American Journal of Criminal Justice 45*(4), 546–562.

Heitzenrater, C. D. and A. C. Simpson (2016). Policy, statistics and questions: Reflections on uk cyber security disclosures. *Journal of Cybersecurity 2*(1), 43–56.

Hota, A. R., A. A. Clements, S. Bagchi, and S. Sundaram (2018). A game-theoretic framework for securing interdependent assets in networks. In *Game theory for security and risk management*, pp. 157–184. Springer.

Kalashnykova, N., V. Kalashnikov, and J. G. F.-M. niz (2021). Bilevel optimization and conjectural equilibrium: Theoretical results and numerical algorithms (an invited tutorial paper). *Journal of Combinatorics, Information & System Sciences 46*, 19–113.

Kemp, S., D. Buil-Gil, F. Miró-Llinares, and N. Lord (2023). When do businesses report cybercrime? findings from a uk study. *Criminology & Criminal Justice 23*(3), 468–489.

Kemp, S., D. Buil-Gil, A. Moneva, F. Miró-Llinares, and N. Díaz-Castaño (2021). Empty streets, busy internet: A time-series analysis of cybercrime and fraud trends during covid-19. *Journal of Contemporary Criminal Justice 37*(4), 480–501.

Klumpp, T., K. A. Konrad, and A. Solomon (2019). The dynamics of majoritarian blotto games. *Games and Economic Behavior 117*, 402–419.

Kovenock, D. and B. Roberson (2012). Conflicts with multiple battlefields. In: Michelle R. Garfinkel Stergios Skaperdas (Eds.), *Oxford handbook of the economics of peace and conflict*.

Kovenock, D. and B. Roberson (2018). The optimal defense of networks of targets. *Economic Inquiry 56*(4), 2195–2211.

Kovenock, D. and B. Roberson (2021). Generalizations of the general lotto and colonel blotto games. *Economic Theory 71*, 997–1032.

Kovenock, D., B. Roberson, and R. M. Sheremeta (2019). The attack and defense of weakest-link networks. *Public Choice 179*(3), 175–194.

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections 24*(3), 43–52.

Levitin, G. and K. Hausken (2010). Resource distribution in multiple attacks against a single target. *Risk Analysis: An International Journal 30*(8), 1231–1239.

Li, X. and J. Zheng (2022). Pure strategy nash equilibrium in 2-contestant generalized lottery colonel blotto games. *Journal of Mathematical Economics 103*, 102771.

Lizzeri, A. and N. Persico (2001). The provision of public goods under alternative electoral incentives. *American Economic Review 91*(1), 225–239.

Marks, M. and R. Croson (1998). Alternate rebate rules in the provision of a threshold public good: An experimental investigation. *Journal of Public Economics 67*, 195–220.

McBride, M. (2006a). Discrete public goods under threshold uncertainty. *Journal of Public Econoimcs 90*, 1181–1199.

McBride, M. (2006b). Imperfect monitoring in communication networks. *Journal of Economic Theory 126*, 97–119.

McBride, M. (2006c). Limited observation in mutual consent networks. *Advances in Theoretical Economics 6*, Article 3.

McBride, M. (2008). Position-specific information in social networks: Are you connected? *Mathematical Social Sciences 56*, 283–295.

McBride, M. (2010). Threshold uncertainty in discrete public good games: An experimental study. *Economics of Governance 11*, 77–99.

McBride, M. and D. Hewitt (2013). The enemy you can't see: An investigation of the disruption of dark networks. *Journal of Economic Behavior & Organization 93*, 32–50.

McCrohan, K. F., K. Engel, and J. W. Harvey (2010). Influence of awareness and training on cyber security. *Journal of internet Commerce 9*(1), 23–41.

Menezes, F., P. Monteiro, and A. Temimi (2001). Private provision of discrete public goods with incomplete information. *Journal of Mathematical Econoimcs 35*, 493–514.

Montero, M., A. Possajennikov, M. Sefton, and T. L. Turocy (2016). Majoritarian blotto contests with asymmetric battlefields: an experiment on apex games. *Economic Theory 61*, 55–89.

Moore, S. (2022). Gartner identifies three factors influencing growth in security spending. URL: `https://www.gartner.com/en/newsroom/press-releases/2022-10-13-gartner-identifies-three-factors-influencing-growth-i`.

Moore, T. (2010). The economics of cybersecurity: Principles and policy options. *International Journal of Critical Infrastructure Protection 3*(3-4), 103–117.

Morselli, C., C. Giguère, and K. Petit (2007). The efficiency/security trade-off in criminal networks. *Social networks 29*(1), 143–153.

NICCS (2024). A glossary of common cybersecurity words and phrases. URL: `https://niccs.cisa.gov/cybersecurity-career-resources/vocabulary#letter-v`.

NIST (2024). Glossary. URL: `https://csrc.nist.gov/glossary`.

Nitzan, S. and R. Romano (1990). Private provision of a discrete public good with uncertain cost. *Journal of Public Economics 42*, 357–370.

Offerman, T., J. Sonnemans, and A. Schram (1996). Value orientations, expectations, and voluntary contributions in public goods. *Economic Journal 106*, 817–845.

Offerman, T., J. Sonnemans, and A. Schram (2001). Expectation formation in step-level public good games. *Economic Inquiry 39*, 250–269.

Palfrey, T. and H. Rosenthal (1984). Participation and the provision of discrete public goods: A strategic analysis. *Journal of Public Economics 24*, 171–193.

Powell, R. (2009). Sequential, nonzero-sum "blotto": Allocating defensive resources prior to attack. *Games and Economic Behavior 67*(2), 611–615.

Roberson, B. and D. Kvasov (2012). The non-constant-sum colonel blotto game. *Economic Theory 51*, 397–433.

Rubinstein, A. and A. Wolinsky (1994). Rationalizable conjectural equilibrium: Between nash and rationalizability. *Games and Economic Behavior 6*, 299–311.

SANS (2024). Glossary of cyber security terms. URL: `https://www.sans.org/security-resources/glossary-of-terms/`.

Skaperdas, S. (1996). Contest success functions. *Economic theory 7*(2), 283–290.

Suleiman, R. (1997). Provision of step-level public goods under uncertainty: A theoretical analysis. *Rationality and Society 9*, 163–187.

Toulas, B. (2022). California medical group data breach impacts 3.3 million patients. *Bleeping Computer*.

Tullock, G. (1980). Efficient rent seeking. jm buchanan, rd tollison, g. tullock, eds., toward a theory of the rent seeking society. *A & M University Press, College Station, TX: Texas*.

Wellman, M. and J. Hu (1998). Conjectural equilibrium in multiagent learning. *Machine Learning 33*, 179–200.

Wit, A. and H. Wilke (1999). Public good provision under environmental and social uncertainty. *European Journal of Social Psychology 28*, 249–256.

Appendix A

Supplementary material for Chapter 1

Lemma 1. *If node i is not in the component containing the high-value node ($i \notin C$), then $\mathcal{A}$ does not allocate resources to attack i ($a_i^* = 0$).*

Proof. Let $i \notin C$. That is, battlefield i is not in the component containing the high-value node (and therefore is also not the high-value node). Then $V_{\mathcal{A},i} = 0$, so $\mathcal{A}$'s payoff function in equation 1.3 is strictly decreasing in a_i. Therefore, in equilibrium, $\mathcal{A}$ does not allocate resources attacking battlefield i ($a_i^* = 0$). □

Lemma 2. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. If $\mathcal{D}$ allocates defensive resources with an infrastructure technology, d_I, and $\mathcal{A}$ allocates attack resources to each node individually, $\bar{a} = (a_1, \ldots, a_n)$), $\mathcal{A}$ will optimally distribute resources uniformly across all nodes in the component containing the high-value node ($a_i^* = a_j^* = a^* \ \forall \ i, j \in C$).*

Proof. For arbitrary battlefields i and j, such that $i \neq j$, the first order conditions of equation

1.3 are:

$$\frac{\partial u_{\mathcal{A}}}{\partial a_i} : \frac{d_I}{(d_I + a_i^*)^2}\left(\prod_{k\neq i}\frac{d_I}{d_I + a_k}\right)v_{\mathcal{A}} - \alpha \leq 0$$

$$\frac{\partial u_{\mathcal{A}}}{\partial a_j} : \frac{d_I}{(d_I + a_j^*)^2}\left(\prod_{k\neq j}\frac{d_I}{d_I + a_k}\right)v_{\mathcal{A}} - \alpha \leq 0$$

$$\implies \frac{\alpha}{v_{\mathcal{A}}} = \frac{d_I}{(d_I + a_i^*)^2}\frac{d_I}{d_I + a_j^*}\prod_{k\notin\{i,j\}}\frac{d_I}{d_I + a_k} = \frac{d_I}{(d_I + a_j^*)^2}\frac{d_I}{d_I + a_i^*}\prod_{k\notin\{i,j\}}\frac{d_I}{d_I + a_k}$$

$$\implies d_I + a_j^* = d_I + a_i^*$$

$$\implies a_i^* = a_j^* \ \ \forall\ i, j$$

Thus, regardless of $\mathcal{D}$'s defensive allocation, d_I, $\mathcal{A}$ allocates attack resources to arbitrary nodes i and j in the component containing the high-value node such that $a_i^* = a_j^*$. This implies $\mathcal{A}$ will uniformly distribute some level of attack resources, a^*, across all nodes in the component containing the high-value node. □

Lemma 3. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. Let $\mathcal{D}$ allocate defensive resources with an infrastructure technology, d_I, and $\mathcal{A}$ allocate attack resources to each node individually, $\bar{a} = (a_1, \ldots, a_n)$). $\mathcal{A}$ does not attack if $\mathcal{D}$ allocates sufficiently high defensive infrastructure $\left(a^* = 0 \ if \ d_I^* \geq \frac{v_{\mathcal{A}}}{\alpha}\right)$. $\mathcal{D}$ does not allocate defensive resources beyond the point at which $\mathcal{A}$ does not attack $\left(d_I^* \leq \frac{v_{\mathcal{A}}}{\alpha} \ if \ a^* = 0\right)$.*

Proof. From Lemma 2, it follows that

$$\frac{\partial u_{\mathcal{A}}}{\partial a_i} : \frac{d_I^{m+1}}{(d_I + a^*)^{m+2}}v_{\mathcal{A}} - \alpha \leq 0$$

Solving for $a^*(d_I)$:

$$a^*(d_I) \geq \left(\frac{v_{\mathcal{A}} d_I^{m+1}}{\alpha}\right)^{1/(m+2)} - d_I$$

However, because $\mathcal{A}$ cannot allocate $a^* < 0$, set $a^* = 0$ when:

$$\left(\frac{v_{\mathcal{A}} d_I^{m+1}}{\alpha}\right)^{1/(m+2)} \leq d_I$$
$$\frac{v_{\mathcal{A}}}{\alpha} \leq d_I$$

If $a^* = 0$, $u_{\mathcal{D}}$ is strictly decreasing in d_I. Therefore, it follows that $\mathcal{D}$ will never have $d_I^* > \frac{v_{\mathcal{A}}}{\alpha}$. □

Proposition 1. *Consider a model of sequential network defense between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$. $\mathcal{D}$ connects $m \in \{0, n-1\}$ low-value nodes to a high-value node and allocates infrastructure defense, d_I, that uniformly protects the network. $\mathcal{A}$ observes $\mathcal{D}$'s choices and allocates attack resources $\bar{a} = (a_1, \ldots, a_n)$ to each node. $\mathcal{D}$ has a weakest-link objective, receives $v_{\mathcal{D}}(1 + m\beta)$ from a successful defense, and has linear costs. $\mathcal{A}$ has a best-shot objective, receives $v_{\mathcal{A}}$ from a successful attack, and has linear costs with the relative price of attack resources to defense resources represented by α. The outcome at each node is determined by a lottery contest success function. Given the exogeneous parameters, there exist two types of subgame perfect equilibrium allocations that can arise:* Attack *or* Deterrence.

1. **Attack:** *$\mathcal{A}$ uniformly allocates a positive level of attack resources across the component containing the high-value node ($a_i^* = a^* > 0 \; \forall \; i \in C$). If $\nu < \frac{m+2}{(m+1)(1+m\beta)}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, $(d_I^* \; , \; a^* \; \forall \; i)$, are:*

$$d_I^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2}, \text{ and}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1\right].$$

2. **Deterrence:** *$\mathcal{D}$ allocates a sufficiently high level of defensive infrastructure such that $\mathcal{A}$ does not allocate resources to attacking $\mathcal{D}$'s network ($a_i^* = a^* = 0 \; \forall \; i$). If $\nu \geq \frac{m+2}{(m+1)(1+m\beta)}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, $(d_I^* \; , \; a^* \; \forall \; i)$,*

are:

$$d_I^* = \frac{v_{\mathcal{A}}}{\alpha}, \text{ and}$$
$$a^* = 0.$$

Proof. From Lemma 3, $\mathcal{A}$'s optimal allocation of attack resources at battlefield i given $\mathcal{D}$'s security infrastructure, d_I, is

$$a^*(d_I) = a_i^*(d_I) = \left(\frac{v_{\mathcal{A}} d_I^{m+1}}{\alpha}\right)^{1/(m+2)} - d_I$$

Therefore, $\mathcal{D}$'s expected payoff as a function of $\mathcal{A}$'s best-response is:

$$\begin{aligned} u_{\mathcal{D}}(a^*(d_I)) &= \left(\prod_{i=1}^{m+1}\left(\frac{d_I}{d_I + a^*(d_I)}\right)\right) v_{\mathcal{D}}(1+m\beta) - d_I \\ &= \left(\frac{d_I}{\left(\frac{v_{\mathcal{A}} d_I^{m+1}}{\alpha}\right)^{1/(m+2)}}\right)^{m+1} v_{\mathcal{D}}(1+m\beta) - d_I \\ &= \left(\frac{d_I}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} v_{\mathcal{D}}(1+m\beta) - d_I. \end{aligned}$$

Solving for d_I^* $\left(\text{letting } \nu = \frac{v_{\mathcal{D}}}{v_{\mathcal{A}}/\alpha}\right)$ gives us:

$$\begin{aligned} \frac{\partial u_{\mathcal{D}}(a^*(d_I))}{\partial d_I} : \left(\frac{\alpha}{v_{\mathcal{A}}}\right)^{(m+1)/(m+2)} \left(\frac{m+1}{m+2}\right) (d_I^*)^{-1/(m+2)} v_{\mathcal{D}}(1+m\beta) - 1 = 0 \\ \implies d_I^* = \left[\left(\frac{\alpha}{v_{\mathcal{A}}}\right)^{(m+1)/(m+2)} \left(\frac{m+1}{m+2}\right) v_{\mathcal{D}}(1+m\beta)\right]^{m+2} \\ = \left(\frac{\alpha}{v_{\mathcal{A}}}\right)^{m+1} v_{\mathcal{D}}^{m+2} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \\ = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2}, \text{ for } m \in \{0, n-1\} \end{aligned}$$

Substituting d_I^* into $a^*(d_I^*)$ to solve for a^* as a function of exogenous parameters:

$$\begin{aligned}
a^*(d_I^*) = a^* &= \left(\frac{(v_{\mathcal{A}})(d_I^*)^{m+1}}{\alpha}\right)^{1/(m+2)} - d_I^* \\
&= d_I^*\left[\left(\frac{v_{\mathcal{A}}}{\alpha d_I^*}\right)^{1/(m+2)} - 1\right] \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \left[\left(\frac{v_{\mathcal{A}}/\alpha}{\nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2}}\right)^{1/(m+2)} - 1\right] \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1\right].
\end{aligned}$$

By Lemma 3, if $d_I^* \geq \frac{v_{\mathcal{A}}}{\alpha}$, then $\mathcal{A}$ sets $a^* = 0$ for all i. Conversely, if $a^* = 0$, $\mathcal{D}$ must have $d_I^* \leq \frac{v_{\mathcal{A}}}{\alpha}$. This implies the subgame perfect equilibrium allocations are $(d_I^*, a^* \; \forall \; i) = \left(\frac{v_{\mathcal{A}}}{\alpha}, 0\right)$ if

$$\begin{aligned}
\nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} &\geq \frac{v_{\mathcal{A}}}{\alpha} \\
\left(\nu\left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right)^{m+2} &\geq 1 \\
\nu &\geq \frac{m+2}{(m+1)(1+m\beta)} \text{ for } m \in \{0, n-1\}.
\end{aligned}$$

To see more explicitly that $(d_I^*, a^* \; \forall \; i) = \left(\frac{v_{\mathcal{A}}}{\alpha}, 0\right)$ is the subgame perfect equilibrium allocation if $\nu \geq \frac{m+2}{(m+1)(1+m\beta)}$, I show that $\mathcal{D}$'s payoff for $d_I = \frac{v_{\mathcal{A}}}{\alpha}$ is greater than for $d_I < \frac{v_{\mathcal{A}}}{\alpha}$:

$$\begin{aligned}
u_{\mathcal{D}}\left(a^*\left(d_I = \frac{v_{\mathcal{A}}}{\alpha}\right)\right) &> u_{\mathcal{D}}\left(a^*\left(d_I < \frac{v_{\mathcal{A}}}{\alpha}\right)\right) \\
\left(\prod_{i=1}^{m+1} 1\right) v_{\mathcal{D}}(1+m\beta) - d_I &> \left(\frac{d_I}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} v_{\mathcal{D}}(1+m\beta) - d_I \\
1 &> \left(\frac{d_I}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)}.
\end{aligned}$$

Note that the inequality in the last line must hold because $d_I < \frac{v_{\mathcal{A}}}{\alpha}$ implies $\frac{d_I}{v_{\mathcal{A}}/\alpha} < 1$. Therefore $\mathcal{D}$ receives a higher payoff from deterring $\mathcal{A}$'s attack by allocating $d_I^* = \frac{v_{\mathcal{A}}}{\alpha}$ when $\nu \geq \frac{m+2}{(m+1)(1+m\beta)}$ for $m \in \{0, n-1\}$.

On the other hand, if $\nu < \frac{m+2}{(m+1)(1+m\beta)}$, the subgame perfect equilibrium allocation is $(d_I^*, a^* \ \forall \ i)$ such that

$$d_I^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1 \right],$$

as solved for earlier in this proof. □

Proposition 2. *Based on the subgame perfect equilibrium allocations in Proposition 1, where $\mathcal{D}$ allocates defensive resources with an infrastructure technology, d_I, the subgame equilibrium profits for $\mathcal{D}$ and $\mathcal{A}$, respectively, are*

$$u_{\mathcal{D}}^* = \begin{cases} (1+m\beta)v_{\mathcal{D}} - \frac{v_{\mathcal{A}}}{\alpha} & if \ d_I^* = \frac{v_{\mathcal{A}}}{\alpha} \\ \left(\frac{1+m\beta}{m+2} \right)^{m+2} (\nu(m+1))^{m+1} v_{\mathcal{D}} & if \ d_I^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

$$u_{\mathcal{A}}^* = \begin{cases} 0 & if \ d_I^* = \frac{v_{\mathcal{A}}}{\alpha} \\ v_{\mathcal{A}} \left\{ 1 - \nu^m \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+1} \left[\nu - (m+1) + (m+1)\nu^2 \left(\frac{(m+1)(1+m\beta)}{m+2} \right) \right] \right\} & if \ d_I^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right), \end{cases}$$

where $d_I^* = \frac{v_{\mathcal{A}}}{\alpha}$ in the "Deterrence" subgame equilibrium and $d_I^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right)$ in the "Attack" subgame equilibrium.

Proof. From Lemma 3, if $d_I^* = v_{\mathcal{A}}/\alpha$ ("Deterrence" equilibrium), then $a^* = 0$ for all $i \in C$.

In this case, $\mathcal{D}$ deters $\mathcal{A}$ from attacking and receives a payoff of

$$\begin{aligned} u_{\mathcal{D}}^* &= \left(\prod_{i=1}^{m+1} \frac{d_I^*}{d_I^* + a^*} \right) v_{\mathcal{D}}(1 + m\beta) - d_I^* \\ &= \left(\prod_{i=1}^{m+1} 1 \right) v_{\mathcal{D}}(1 + m\beta) - \frac{v_{\mathcal{A}}}{\alpha} \\ &= v_{\mathcal{D}}(1 + m\beta) - \frac{v_{\mathcal{A}}}{\alpha} \end{aligned}$$

such that $m \in \{0, n-1\}$. $u_{\mathcal{A}}^* = 0$ because $\mathcal{A}$ neither wins the contest nor spends resources.

From Proposition 1, if $d_I^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right)$ ("Attack" equilibrium), then it follows the equilibrium allocations are:

$$\begin{aligned} d_I^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \\ a^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(m+1)(1+m\beta)} - 1 \right] \end{aligned}$$

Using these equilibrium allocations, I now calculate $\mathcal{D}$'s equilibrium payoff as:

$$
\begin{aligned}
u^*_{\mathcal{D}} &= \left(\prod_{i=1}^{m+1} \frac{d^*_I}{d^*_I + a^*(d^*_I)}\right) v_{\mathcal{D}}(1+m\beta) - d^*_I \\
&= \left(\prod_{i=1}^{m+1} \frac{d^*_I}{d^*_I + \left(\dfrac{v_{\mathcal{A}} d_I^{m+1}}{\alpha}\right)^{1/(m+2)}} - d^*_I\right) v_{\mathcal{D}}(1+m\beta) - \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \\
&= \left(\prod_{i=1}^{m+1} \left(\frac{d^*_I}{v_{\mathcal{A}}/\alpha}\right)^{1/(m+2)}\right) v_{\mathcal{D}}(1+m\beta) - \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \\
&= \left(\frac{\nu^{m+1} v_{\mathcal{D}} \left(\dfrac{(m+1)(1+m\beta)}{m+2}\right)^{m+2}}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} v_{\mathcal{D}}(1+m\beta) - \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \\
&= \left(\nu \left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right)^{m+1} v_{\mathcal{D}}(1+m\beta) - \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+2} \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+1} \left[(1+m\beta) - \frac{(m+1)(1+m\beta)}{m+2}\right] \\
&= \nu^{m+1} v_{\mathcal{D}} (m+1)^{m+1} \left(\frac{1+m\beta}{m+2}\right)^{m+2}
\end{aligned}
$$

and $\mathcal{A}$'s equilibrium payoff as:

$$
\begin{aligned}
u^*_{\mathcal{A}} &= \left(1 - \prod_{i=1}^{m+1} \frac{d^*_I}{d^*_I + a^*(d^*_I)}\right) v_{\mathcal{A}} - \sum_{i=1}^{m+1} \alpha a^* \\
&= \left(1 - \left(\nu \left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right)^{m+1}\right) v_{\mathcal{A}} \\
&\quad - (m+1)\alpha\nu^{m+1} v_{\mathcal{D}} \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+1} \left[\nu^{m/(m+2)} - \frac{(m+1)(1+m\beta)}{m+2}\right] \\
&= v_{\mathcal{A}} \Bigg\{ 1 - \left(\nu \left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right)^{m+1} - (m+1)\nu^m \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+1} \\
&\quad + (m+1) \left(\nu \left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right)^{m+2} \Bigg\} \\
&= v_{\mathcal{A}} \left\{ 1 - \nu^m \left(\frac{(m+1)(1+m\beta)}{m+2}\right)^{m+1} \left[\nu - (m+1) + (m+1)\nu^2 \left(\frac{(m+1)(1+m\beta)}{m+2}\right)\right] \right\}
\end{aligned}
$$

Thus, I have found the subgame equilibrium payoffs for $\mathcal{D}$ and $\mathcal{A}$ in both the "Deterrence" and "Attack" equilibria. $\square$

Lemma 4. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. If both $\mathcal{D}$ and $\mathcal{A}$ allocate resources to each node individually ($\bar{d} = (d_1, \ldots, d_n)$ and $\bar{a} = (a_1, \ldots, a_n)$), both $\mathcal{D}$ and $\mathcal{A}$ will optimally distribute resources uniformly across all nodes in the component containing the high-value node ($d_i^* = d_j^* = d^*$ and $a_i^* = a_j^* = a^* \ \forall \ i, j \in C$).*

Proof. First, solving for $a_i^*(\bar{d})$ and finding the relationship with $a_j^*(\bar{d})$:

$$\frac{\partial u_{\mathcal{A}}}{\partial a_i} : \frac{d_i}{(d_i + a_i^*(\bar{d}))^2} \left(\prod_{k \neq i} \frac{d_k}{d_k + a_k} \right) v_{\mathcal{A}} - \alpha = 0$$

$$(d_i + a_i^*(\bar{d}))^2 = \frac{d_i}{\alpha} \left(\prod_{k \neq i} \frac{d_k}{d_k + a_k} \right) v_{\mathcal{A}}$$

$$a_i^*(\bar{d}) = \left(\frac{d_i v_{\mathcal{A}}}{\alpha} \prod_{k \neq i} \frac{d_k}{d_k + a_k} \right)^{0.5} - d_i$$

$$\frac{\partial u_{\mathcal{A}}}{\partial a_j} : \frac{d_j}{(d_j + a_j^*(\bar{d}))^2} \left(\prod_{k \neq j} \frac{d_k}{d_k + a_k} \right) v_{\mathcal{A}} - \alpha = 0$$

$$\frac{\alpha}{v_{\mathcal{A}}} = \frac{d_i}{(d_i + a_i^*(\bar{d}))^2} \frac{d_j}{d_j + a_j^*(\bar{d})} \prod_{k \notin \{i,j\}} \frac{d_k}{d_k + a_k} = \frac{d_i}{d_i + a_i^*(\bar{d})} \frac{d_j}{(d_j + a_j^*(\bar{d}))^2} \prod_{k \notin \{i,j\}} \frac{d_k}{d_k + a_k}$$

$$d_i + a_i^*(\bar{d}) = d_j + a_j^*(\bar{d})$$

$$a_j^*(\bar{d}) = d_i + a_i^*(\bar{d}) - d_j \ \ \forall \ j \neq i$$

$$\implies \frac{d_i}{(d_i + a_i^*(\bar{d}))^2} \left(\prod_{k \neq i} \frac{d_k}{d_k + d_i + a_i^*(\bar{d}) - d_k} \right) v_{\mathcal{A}} - \alpha = 0$$

$$\frac{d_i v_{\mathcal{A}}}{\alpha (d_i + a_i^*(\bar{d}))^{m+2}} \prod_{k \neq i} d_k = 1$$

$$a_i^*(\bar{d}) = \left(\frac{v_{\mathcal{A}}}{\alpha} \prod_{i=1}^{m+1} d_i \right)^{1/(m+2)} - d_i \ \ \forall \ i$$

Substituting $\bar{a}^* = (a_1^*(\bar{d}), \ldots, a_n^*(\bar{d}))$ into $u_{\mathcal{D}}$ reduces to the following function:

$$\begin{aligned}
u_{\mathcal{D}}(\bar{a}^*) &= \left(\prod_{i=1}^{m+1} \frac{d_i}{d_i + a_i^*(\bar{d})} \right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i+1}^{m+1} d_i \\
&= \left(\prod_{i=1}^{m+1} \frac{d_i}{\left(\frac{v_{\mathcal{A}}}{\alpha} \prod_{i=1}^{m+1} d_i \right)^{1/(m+2)}} \right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i+1}^{m+1} d_i \\
&= \left(\frac{1}{v_{\mathcal{A}}/\alpha} \right)^{(m+1)/(m+2)} \left(\prod_{i=1}^{m+1} \frac{d_i^{(m+1)/(m+2)}}{\left(\prod_{j \neq i} d_j \right)^{1/(m+2)}} \right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i+1}^{m+1} d_i \\
&= \left(\frac{\prod_{i=1}^{m+1} d_i^{(m+1)/(m+2)}}{\prod_{i=1}^{m+1} d_i^{m/(m+2)}} \right) (\nu^{m+1} v_{\mathcal{D}})^{1/(m+2)} (1 + m\beta) - \sum_{i+1}^{m+1} d_i \\
&= \left(\prod_{i=1}^{m+1} d_i^{1/(m+2)} \right) (\nu^{m+1} v_{\mathcal{D}})^{1/(m+2)} (1 + m\beta) - \sum_{i+1}^{m+1} d_i.
\end{aligned}$$

From this, we can solve for $\mathcal{D}$'s optimal defensive allocation at nodes d_i^* and d_j^* to see that $d_i^* = d_j^*$ for arbitrary i and j such that $i \neq j$.

$$\begin{aligned}
\frac{\partial u_{\mathcal{D}}(\bar{a}^*)}{\partial d_i} &: \left(\prod_{k \neq i} d_k^{1/(m+2)} \right) d_i^{-(m+1)/(m+2)} (\nu^{m+1} v_{\mathcal{D}})^{1/(m+2)} \frac{(1 + m\beta)}{m+2} - 1 = 0 \\
\frac{\partial u_{\mathcal{D}}(\bar{a}^*)}{\partial d_j} &: \left(\prod_{k \neq j} d_k^{1/(m+2)} \right) d_j^{-(m+1)/(m+2)} (\nu^{m+1} v_{\mathcal{D}})^{1/(m+2)} \frac{(1 + m\beta)}{m+2} - 1 = 0 \\
&\implies \frac{d_i^{(m+1)/(m+2)}}{d_j^{1/(m+2)}} = \frac{d_j^{(m+1)/(m+2)}}{d_i^{1/(m+2)}} \\
&\quad d_i^{m/(m+2)} = d_j^{m/(m+2)} \\
&\quad d_i = d_j \quad \forall\ i, j \ \text{ s.t. } i \neq j
\end{aligned}$$

From before, $\mathcal{A}$'s optimal best-response attack allocation at arbitrary nodes i and j to $\mathcal{D}$'s defensive allocation is $a_j^*(\bar{d}) = d_i + a_i^*(\bar{d}) - d_j$. Because $\mathcal{D}$ has $d_i = d_j$ at equilibrium, though, this implies $a_j = i^*(\bar{d}) = a_j^*(\bar{d})$ for arbitrary nodes i and j. Therefore, both $\mathcal{D}$ and $\mathcal{A}$ will uniformly distribute some level of resources, d^* and a^* respectively, across all nodes in the component containing the high-value node. □

Lemma 5. *Consider a sequential network defense contest between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$, where victory is determined by a weakest-link versus best-shot lottery contest. Let both $\mathcal{D}$ and $\mathcal{A}$ allocate resources to each node individually ($\bar{d} = (d_1, \ldots, d_n)$ and $\bar{a} = (a_1, \ldots, a_n)$). $\mathcal{A}$ will not attack a node if $\mathcal{D}$ allocates sufficiently high defensive resources to that node $\left(a_i^* = 0 \; if \; d_i^* \geq \frac{v_{\mathcal{A}}}{\alpha}\right)$. $\mathcal{D}$ does not allocate defensive resources beyond the point at which $\mathcal{A}$ does not attack a node $\left(d_i^* \leq \frac{v_{\mathcal{A}}}{\alpha} \; if \; a_i^* = 0\right)$.*

Proof. Solving first for $a_i^*(\bar{d})$ for all i:

$$\begin{aligned} & \frac{\partial u_{\mathcal{A}}}{\partial a_i} : \frac{d_i}{(d_i + a_i^*(\bar{d}))^2} \prod_{j \neq i} \frac{d_j}{d_j + a_j} v_{\mathcal{A}} - \alpha = 0 \\ \text{Lemma 4} \quad & \implies \frac{d_i^{m+1}}{(d_i + a_i^*(\bar{d}))^{m+2}} v_{\mathcal{A}} = \alpha \\ & \implies a_i^*(\bar{d}) = \left(\frac{v_{\mathcal{A}} d_i^{m+1}}{\alpha}\right)^{1/(m+2)} - d_i \quad \forall \; i \end{aligned}$$

However, because $\mathcal{A}$ cannot allocate $a_i^* < 0$, set $a_i^* = 0$ when:

$$\begin{aligned} \left(\frac{v_{\mathcal{A}} d_i^{m+1}}{\alpha}\right)^{1/(m+2)} & \leq d_i \\ \frac{v_{\mathcal{A}}}{\alpha} & \leq d_i \quad \forall \; i. \end{aligned}$$

If $a_i^* = 0$, $u_{\mathcal{D}}$ is strictly decreasing in d_i. Therefore, it follows that $\mathcal{D}$ will never have $d_i^* > \frac{v_{\mathcal{A}}}{\alpha}$. □

Proposition 3. *Consider a model of sequential network defense between a defender, $\mathcal{D}$, and an attacker, $\mathcal{A}$. $\mathcal{D}$ connects $m \in \{0, n-1\}$ low-value nodes to a high-value node and allocates defensive resources $\bar{d} = (d_1, \ldots, d_n)$ to each node. $\mathcal{A}$ observes $\mathcal{D}$'s choices and allocates attack resources $\bar{a} = (a_1, \ldots, a_n)$ to each node. $\mathcal{D}$ has a weakest-link objective, receives $v_{\mathcal{D}}(1 + m\beta)$ from a successful defense, and has linear costs. $\mathcal{A}$ has a best-shot objective, receives $v_{\mathcal{A}}$ from a successful attack, and has linear costs with the relative price of attack resources to defense*

resources represented by α. The outcome at each node is determined by a lottery contest success function. Given the exogeneous parameters, there exist two types of subgame perfect equilibrium allocations that can arise: Attack *or* Deterrence.

1. **Attack:** *$\mathcal{D}$ and $\mathcal{A}$ each allocate a uniform and positive level of resources to nodes in the component containing the high-value node ($d_i^* = d^* \ \forall \ i \in C$ and $a_i^* = a^* \ \forall \ i \in C$). If $\nu < \dfrac{m+2}{1+m\beta}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, ($d^* \ \forall \ i \ , \ a^* \ \forall \ i$), are:*

$$d^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2} \right)^{m+2}, \ \textit{and}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(1+m\beta)} - 1 \right]$$

2. **Deterrence:** *$\mathcal{D}$ allocates a sufficiently high level of defensive resources to each node in the component containing the high-value node such that $\mathcal{A}$ does not allocate resources to attacking $\mathcal{D}$'s network*
($a_i^* = a^* = 0 \ \forall \ i$). If $\nu \geq \dfrac{m+2}{1+m\beta}$, $\mathcal{D}$ and $\mathcal{A}$'s subgame perfect equilibrium allocations, ($d^* \ \forall \ i \ , \ a^* \ \forall \ i$), are:

$$d^* = \frac{v_{\mathcal{A}}}{\alpha}, \ \textit{and}$$
$$a^* = 0$$

Proof. I solve first for $a_i^*(\bar{d})$ for all nodes in the component containing the high-value node ($i \in C$).

$$\text{Lemma 4} \implies a^*(\bar{d^*}) = a_i^*(\bar{d^*}) = \left(\frac{v_{\mathcal{A}} (d_i^*)^{m+1}}{\alpha} \right)^{1/(m+2)} - d_i^* \ \ \forall \ i$$

Similarly,

$$\frac{\partial u_{\mathcal{D}}(\bar{a}^*)}{\partial d_i} : \left(\prod_{k\neq i} d_k^{1/(m+2)}\right)(d_i^*)^{-(m+1)/(m+2)}(\nu^{m+1}v_{\mathcal{D}})^{1/(m+2)}\frac{(1+m\beta)}{m+2} - 1 = 0$$

$$\text{Lemma 4} \implies (d_i^*)^{-1/(m+2)}(\nu^{m+1}v_{\mathcal{D}})^{1/(m+2)}\frac{(1+m\beta)}{m+2} - 1 = 0$$

$$d^* = d_i^* = \nu^{m+1}v_{\mathcal{D}}\left(\frac{(1+m\beta)}{m+2}\right)^{m+2} \quad \forall\ i.$$

Now I can solve for $a^*(\bar{d^*})$ as a function of exogenous parameters.

$$\begin{aligned}
a^*(\bar{d^*}) &= \left(\frac{(v_{\mathcal{A}})(d^*)^{m+1}}{\alpha}\right)^{1/(m+2)} - d^* \\
&= d^*\left[\left(\frac{v_{\mathcal{A}}}{\alpha d^*}\right)^{1/(m+2)} - 1\right] \\
&= \nu^{m+1}v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2}\left[\left(\frac{v_{\mathcal{A}}}{\alpha\nu^{m+1}v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2}}\right)^{1/(m+2)} - 1\right] \\
&= \nu^{m+1}v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2}\left[\frac{m+2}{\nu(1+m\beta)} - 1\right]
\end{aligned}$$

By Lemma 5, if $d^* \geq \frac{v_{\mathcal{A}}}{\alpha}$, then $\mathcal{A}$ sets $a^* = 0$ for all i. Conversely, if $a^* = 0$, $\mathcal{D}$ must have $d^* \leq \frac{v_{\mathcal{A}}}{\alpha}$. This implies the subgame perfect equilibrium allocations are $(d^*\ \forall\ i, a^*\ \forall\ i) = \left(\frac{v_{\mathcal{A}}}{\alpha}, 0\right)$ if:

$$\begin{aligned}
\nu^{m+1}v_{\mathcal{D}}\left(\frac{1+m\beta}{m+2}\right)^{m+2} &\geq \frac{v_{\mathcal{A}}}{\alpha} \\
\left(\nu\left(\frac{1+m\beta}{m+2}\right)\right)^{m+2} &\geq 1 \\
\nu &\geq \frac{m+2}{1+m\beta} \text{ for } m \in \{0, n-1\}
\end{aligned}$$

To see more explicitly that $(d^*\ \forall\ i, a^*\ \forall\ i) = \left(\frac{v_{\mathcal{A}}}{\alpha}, 0\right)$ is the subgame perfect equilibrium

allocation if $\nu \geq \frac{m+2}{1+m\beta}$, I show that $\mathcal{D}$'s payoff from $d^* = \frac{v_{\mathcal{A}}}{\alpha}$ is greater than for $d < \frac{v_{\mathcal{A}}}{\alpha}$:

$$u_{\mathcal{D}}\left(a^*\left(d^* = \frac{v_{\mathcal{A}}}{\alpha}\ \forall\ i\right)\right) > u_{\mathcal{D}}\left(a^*\left(d < \frac{v_{\mathcal{A}}}{\alpha}\ \forall\ i\right)\right)$$

$$\left(\prod_{i=1}^{m+1} 1\right) v_{\mathcal{D}}(1+m\beta) - \sum_{i=1}^{m+1} \frac{v_{\mathcal{A}}}{\alpha} > \left(\prod_{i=1}^{m+1} \left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{1/(m+2)}\right) v_{\mathcal{D}}(1+m\beta) - \sum_{i=1}^{m+1} d$$

$$1 - \frac{m+1}{\nu(1+m\beta)} > \left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} - \frac{m+1}{v_{\mathcal{D}}(1+m\beta)} d$$

$$1 > \left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} + \frac{m+1}{\nu(1+m\beta)}\left(1 - \frac{d}{v_{\mathcal{A}}/\alpha}\right).$$

Notice that the right hand side of the above inequality is increasing as ν increases. Furthermore, the deterrence equilibrium occurs if $\nu \geq \frac{m+2}{1+m\beta}$, so the right hand side is maximal with respect to ν when $\nu = \frac{m+2}{1+m\beta}$. Therefore, the following inequality necessarily implies the above inequality:

$$1 > \left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} + \frac{m+1}{m+2}\left(1 - \frac{d}{v_{\mathcal{A}}/\alpha}\right)$$

$$1 > (m+2)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} - (m+1)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right).$$

Now I can show that the right hand side is increasing as d increases:

$$\frac{\partial RHS}{\partial d} = \left(\frac{1}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} (m+1) d^{-1/(m+2)} - \frac{m+1}{v_{\mathcal{A}}/\alpha}$$

$$= \frac{m+1}{v_{\mathcal{A}}/\alpha}\left[\left(\frac{v_{\mathcal{A}}/\alpha}{d}\right)^{1/(m+2)} - 1\right] > 0 \quad \text{because} \quad d < \frac{v_{\mathcal{A}}}{\alpha}.$$

Continuing to look at the right hand side of the inequality and using the open upper bound

of $d = \frac{v_{\mathcal{A}}}{\alpha}$, this implies

$$(m+2)(1)^{(m+1)/(m+2)} - (m+1)(1) > (m+2)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} - (m+1)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)$$
$$1 > (m+2)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} - (m+1)\left(\frac{d}{v_{\mathcal{A}}/\alpha}\right) \quad \forall\ d < \frac{v_{\mathcal{A}}}{\alpha}.$$

Thus, this shows $\mathcal{D}$ does indeed receive a higher payoff from deterring $\mathcal{A}$'s attack by allocating $d^* = \frac{v_{\mathcal{A}}}{\alpha}$ for all nodes i in the component containing the high-value node when $\nu \geq \frac{m+2}{1+m\beta}$ for $m \in \{0, n-1\}$.

On the other hand, if $\nu < \frac{m+2}{1+m\beta}$, the subgame perfect equilibrium allocation is $(d^*\ \forall\ i, a^*\ \forall\ i)$ such that

$$d^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2}$$
$$a^* = \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2} \left[\frac{m+2}{\nu(1+m\beta)} - 1\right]$$

as solved for earlier in the proof. □

Proposition 4. *Based on the subgame perfect equilibrium allocations in Proposition 3, where $\mathcal{D}$ allocates defensive resources to each node individually, $\bar{d} = (d_1, \ldots, d_n)$, the subgame equilibrium profits for $\mathcal{D}$ and $\mathcal{A}$, respectively, are*

$$u^*_{\mathcal{D}} = \begin{cases} v_{\mathcal{D}}(1+m\beta) - (m+1)\frac{v_{\mathcal{A}}}{\alpha} & if\ d^* = \frac{v_{\mathcal{A}}}{\alpha} \\ \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2} & if\ d^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

$$u^*_{\mathcal{A}} = \begin{cases} 0 & if\ d^* = \frac{v_{\mathcal{A}}}{\alpha} \\ v_{\mathcal{A}}\left[1 - \nu^m \left(\frac{1+m\beta}{m+2}\right)^{m+1} \left(\nu + m + 1 - (m+1)\nu^2 \left(\frac{1+m\beta}{m+2}\right)\right)\right] & if\ d^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right) \end{cases}$$

where $d^* = \frac{v_{\mathcal{A}}}{\alpha}\ \forall\ i \in C$ in the "Deterrence" subgame equilibrium and $d^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right)\ \forall\ i \in C$ in the "Attack" subgame equilibrium.

Proof. From Lemma 5, if $d^* = v_{\mathcal{A}}/\alpha$ for all $i \in C$ ("Deterrence" equilibrium), $a^* = 0$ for all $i \in C$ (where C is the network containing the high-value node). In this case, $\mathcal{D}$ deters $\mathcal{A}$ from attacking and receives a payoff of

$$\begin{aligned} u^*_{\mathcal{D}} &= \left(\prod_{i=1}^{m+1} \frac{d^*_i}{d^*_i + a^*_i} \right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i=1}^{m+1} d^*_i \\ &= \left(\prod_{i=1}^{m+1} 1 \right) v_{\mathcal{D}}(1 + m\beta) - \sum_{i=1}^{m+1} \frac{v_{\mathcal{A}}}{\alpha} \\ &= v_{\mathcal{D}}(1 + m\beta) - (m+1)\frac{v_{\mathcal{A}}}{\alpha} \end{aligned}$$

such that $m \in \{0, n-1\}$. Because $\mathcal{A}$ does not expend any resources attacking the network, it is clear she has a cost of zero and receives a benefit of zero. That is, $u^*_{\mathcal{A}} = 0$.

From Proposition 3, if $d^* \in \left(0, \frac{v_{\mathcal{A}}}{\alpha}\right)$ for all $i \in C$ ("Attack" equilibrium), it follows that the equilibrium allocations are:

$$\begin{aligned} d^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{1 + m\beta}{m+2} \right)^{m+2} \quad \forall\ i \text{ and} \\ a^* &= \nu^{m+1} v_{\mathcal{D}} \left(\frac{1 + m\beta}{m+2} \right)^{m+2} \left[\frac{m+2}{\nu(1 + m\beta)} - 1 \right] \quad \forall\ i. \end{aligned}$$

Using these equilibrium allocations, I now calculate $\mathcal{D}$'s equilibrium payoff:

$$\begin{aligned}
u^*_{\mathcal{D}} &= \left(\prod_{i=1}^{m+1} \frac{d_i^*}{d_i^* + a_i^*}\right) v_{\mathcal{D}}(1+m\beta) - \sum_{i=1}^{m+1} d_i^* \\
&= \left(\prod_{i=1}^{m+1} \frac{d_i^*}{d_i^* + \left(\dfrac{(v_{\mathcal{A}})(d_i^*)^{m+1}}{\alpha}\right)^{1/(m+2)} - d_i^*}\right) v_{\mathcal{D}}(1+m\beta) - \sum_{i=1}^{m+1} d_i^* \\
&= \left(\prod_{i=1}^{m+1} \left(\frac{d_i^*}{v_{\mathcal{A}}/\alpha}\right)^{1/(m+2)}\right) v_{\mathcal{D}}(1+m\beta) - \sum_{i=1}^{m+1} d_i^* \\
&= \left(\frac{\nu^{m+1} v_{\mathcal{D}} \left(\dfrac{1+m\beta}{m+2}\right)^{m+2}}{v_{\mathcal{A}}/\alpha}\right)^{(m+1)/(m+2)} v_{\mathcal{D}}(1+m\beta) - (m+1)\nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2} \\
&= \left(\nu\left(\frac{1+m\beta}{m+2}\right)\right)^{m+1} v_{\mathcal{D}}(1+m\beta) - (m+1)\nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2} \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+1} \left[(1+m\beta) - (m+1)\left(\frac{1+m\beta}{m+2}\right)\right] \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+1} \left(\frac{(m+2)(1+m\beta) - (m+1)(1+m\beta)}{m+2}\right) \\
&= \nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+2}
\end{aligned}$$

and $\mathcal{A}$'s equilibrium payoff:

$$\begin{aligned}
u^*_{\mathcal{A}} &= \left(1 - \prod_{i=1}^{m+1} \left(\frac{d_i^*}{d_i^* + a_i^*}\right)\right) v_{\mathcal{A}} - \sum_{i=1}^{m+1} \alpha a_i^* \\
&= \left(1 - \left(\nu\left(\frac{1+m\beta}{m+2}\right)\right)^{m+1}\right) v_{\mathcal{A}} - (m+1)\alpha\nu^{m+1} v_{\mathcal{D}} \left(\frac{1+m\beta}{m+2}\right)^{m+1} \left[\nu^{m/(m+2)} - \left(\frac{1+m\beta}{m+2}\right)\right] \\
&= v_{\mathcal{A}} \left[1 - \left(\nu\left(\frac{1+m\beta}{m+2}\right)\right)^{m+1} - (m+1)\nu^m \left(\frac{1+m\beta}{m+2}\right)^{m+1} + (m+1)\nu^{m+2}\left(\frac{1+m\beta}{m+2}\right)^{m+2}\right] \\
&= v_{\mathcal{A}} \left[1 - \nu^m \left(\frac{1+m\beta}{m+2}\right)^{m+1} \left(\nu + (m+1) - (m+1)\nu^2\left(\frac{1+m\beta}{m+2}\right)\right)\right]
\end{aligned}$$

Thus, I have found the subgame equilibrium payoffs for $\mathcal{D}$ and $\mathcal{A}$ in both the "Deterrence"

and “Attack” equilibria. □

Appendix B

Supplementary material for Chapter 2

Proposition 1 In the threshold public good game with $v > 1$:

(a) The set of pure Nash Equilibria includes the no-contribution and perfect-provision strategy profiles.

(b) The set of pure Nash Equilibria is the same for all levels of feedback.

Proof (a) First consider $s = (0, 0, 0)$ so there are no contributors. With no other players contributing, player i receives $u_i = 0$ by not contributing and $u_i = -1$ by contributing, so not contributing is the unique best response. This holds for all i, so $s = (0, 0, 0)$ is a Nash Equilibrium.

Next, without loss of generality, consider $s = (s_i, s_j, s_k) = (1, 1, 0)$ so there are two contributors. Given s, player i receives $u_i = v - 1 > 0$ by contributing and $u_i = 0$ by not contributing, so contributing is the unique best response. The same is true for j. Player k receives $u_i = v - 1 > 0$ by contributing and $u_i = v$ by not contributing, so not contributing is the unique best response. Thus, $s = (1, 1, 0)$ is a Nash Equilibrium.

Now consider $s = (1, 0, 0)$ with exactly one contributor. Player i receives $u_i = -1$ by contributing and $u_i = 0$ by not contributing, so not contributing is the unique best response. Thus, $s = (1, 0, 0)$ with exactly one contributor is not a Nash Equilibrium.

Finally consider $s = (1, 1, 1)$ with three contributors. Player k receives $u_i = v - 1$ by contributing and $u_i = v$ by not contributing, so not contributing is the unique best response. Thus, $s = (1, 1, 1)$ is not a Nash Equilibrium.

(b) The proof in (a) did not rely on the level of feedback, so it applies for all three feedback conditions. □

Proposition 2 In the threshold public good game with $v > 1$:

(a) Under Full feedback, the set of Conjectural Equilibria is equivalent to the set of Nash Equilibria, i.e., it consists of the no-contribution and perfect-provision strategy profiles.

(b) Under Partial feedback, the set of Conjectural Equilibria includes the Nash Equilibria and the three-contributor strategy profile.

(c) Under None feedback, the set of Conjectural Equilibria includes the Nash Equilibria, the three-contributor strategy profile, and the one-contributor strategy profile.

Proof (a) Because any Nash Equilibrium profile is a Conjectural Equilibrium in which individuals have correct beliefs, we prove the claim by showing that the three-contributor and one-contributor profiles are not Conjectural Equilibria under Full feedback.

Suppose there are three contributors. By i's information partition, i must have beliefs $\pi_i^*(1, 1, 1) = 1$ and $\pi_i^*(s') = 0$ for all $s \neq (1, 1, 1)$. With these beliefs, i's unique best response is to not contribute, but this implies that having three contributors cannot be a Conjectural Equilibrium.

Now suppose there is one contributor. By i's information partition, i must have beliefs $\pi_i^*(1,0,0) = 1$ and $\pi_i^*(s') = 0$ for all $s \neq (1,0,0)$. With these beliefs, i's unique best response is to not contribute, but this implies that having one contributor cannot be a Conjectural Equilibrium.

(b) With Partial feedback, the no-contribution and perfect-provision strategy profiles are still correct-belief Conjectural Equilibria.

Now suppose i is the only contributor, i.e, $s = (1,0,0)$. Observe that $P_i((1,0,0))$ consists of only one strategy profile, i.e., just $(1,0,0)$. If we suppose that $s = (1,0,0)$ is a Conjectural Equilibrium strategy profile, then by conditions (ii-a) and (ii-b) in definition (2.4) it must be the case that $\pi_i^*(1,0,0) = 1$ and $\pi_i^*(s') = 0$ for all $s' \neq (1,0,0)$. In short, i's beliefs must be correct. However, with these beliefs, player i is strictly better off in expectation by not contributing, a contradiction that means there cannot be a Conjectural Equilibrium with exactly one contributor.

Finally suppose there are three contributors, i.e., $s = (1,1,1)$, and observe that $P_i((1,1,1))$ contains three strategy profiles: $(1,1,0)$, $(1,0,1)$, and $(1,1,1)$. According to conditions (ii-a) and (ii-b) in definition (2.4) player i must distribute her belief probability only among those three strategy profiles in a Conjectural Equilibrium. Notice that $\pi_i^*(1,1,0) = 1$ and $\pi_i^*(s) = 0$ for all $s \neq (1,1,0)$ satisfy these conditions for player i. Without loss of generality, construct similar belief for j and k. Then no player's beliefs are contradicted by their feedback, and each player's contribution is a best response to their belief that they are pivotal. Thus, this over-contribution strategy profile is a Conjectural Equilibrium with appropriately incorrect beliefs.

(c) With None feedback, for any pure strategy profile there exists a profile of beliefs that combines with that strategy profile to constitute a Conjectural Equilibrium. For example, if i contributes, then $\pi_i^*(1,1,0) = 1$ and $\pi_i^*(s') = 0$ for all $s' \neq (1,1,0)$ makes contributing

a best response, and if i does not contribute, then $\pi_i^*(0,0,0) = 1$ and $\pi_i^*(s') = 0$ for all $s' \neq (0,0,0)$ makes not contributing a best response. $\square$

Appendix C

Supplementary material for Chapter 3

C.1 Full Descriptive Statistics of Cybersecurity Measures

Table C.1: Cybersecurity Measures Full

Has measure in 2021	No	No	Yes	Yes	Other		
Has measure in 2022	No	Yes	No	Yes	Other	2021	2022
General Protection							
Training in past 12 months	184	108	50	314	18	54.5%	63.2%
AI or ML tools	314	51	40	82	187	20.2%	22.7%
Adherence to at least one standard/accreditation[1]	346	88	45	192	3	35.3%	41.5%
Has some type of cyber insurance	229	59	30	117	239	63.1%	67.7%
Rules and Policies							
Policy to apply software security updates	84	72	70	346	102	65.9%	67.1%
Any monitoring of user activity	119	77	81	349	48	65.6%	66.3%
Rules for storing and moving files containing personal data	22	45	57	513	37	87.2%	85.0%
Backing up data securely via a cloud service	85	61	46	453	29	74.9%	78.6%
Backing up data securely via other means	103	60	74	384	53	70.6%	67.7%
Up-to-date malware protection across all devices	0	10	8	634	22	96.7%	97.0%
Firewalls covering IT network and individual devices	8	16	21	609	20	94.5%	94.7%
Restricting IT admin and access rights	3	7	9	646	9	97.6%	97.5%
Security controls on organisation's devices	2	25	30	592	25	94.2%	93%
All five Cyber Essentials[2]	157	112	101	304	0	60.1%	61.7%
Policy to not pay ransomware	68	70	56	196	284	43.3%	47.3%
Incident Management							
Business Continuity Plan	72	53	42	436	71	74.3%	75.5%
Risk register	121	61	61	323	108	61.3%	61.3%
Documentation on acceptable cyber risk level	271	72	61	111	159	28.0%	31.9%
Documentation identifying most critical assets to protect	111	81	78	304	100	60.2%	61.1%
Written list of IT estate and vulnerabilities	103	82	83	290	116	59.8%	59.5%
Incident Response Plan	164	90	54	298	68	54.2%	60.2%
Held exercise to test cyber incident response	109	52	35	77	401	20.5%	23.1%
Vulnerability Identification							
Vulnerability audit	157	92	79	261	85	53.0%	56.1%
Risk assessment	81	70	69	399	55	71.5%	72.3%
Invested in threat intelligence	242	74	87	123	148	34.4%	33.7%
Used tools for security monitoring	89	66	67	353	99	66.3%	67.2%
Formally assessed risks presented by any partners	326	70	56	106	116	26.0%	29.5%
Visibility							
Staff can access network/files through personal devices	259	84	114	207	10	48.2%	43.3%
Staff can connect to network/files outside workplace	127	47	62	388	50	68.5%	67.1%
Has a VPN for staff connecting remotely	119	45	72	420	18	73.9%	69.7%
Uses a cloud server that stores data/files	93	84	69	410	18	71.8%	74.2%
Uses a physical server that stores data/files	104	27	52	479	12	79.5%	75.5%

[1]ISO 27001, Cyber Essentials standard, and Cyber Essentials Plus standard
[2]Firewalls, secure configurations, access controls, malware protection, and patch management

C.2 Board Involvement

Table C.2 shows how organizations' boards are involved in cybersecurity. Board involvement is an important aspect of cybersecurity as it reflects an organization's cybersecurity culture and overall concern toward addressing relevant risks and vulnerabilities. Panel (a) highlights

specific ways an organization's board is involved in cybersecurity, whereas Panel (b) presents how the board discusses and engages with cybersecurity more generally. Similar to Table 3.2, I highlight how an organization's board changes their involvement in cybersecurity between 2021 and 2022 and provide the overall distribution of organizations with each board characteristic in 2021 and 2022 respectively. Organizations generally seem to have increased board involvement in 2022 relative to 2021, as every measure in Panel (a) has increased adoption and Panel (b) shows there is a seemingly higher degree of cybersecurity conversation and integration in 2022.

Table C.2: Board Involvement in Cybersecurity

(a) **General Board Involvement**

Has measure in 2021 Has measure in 2021	No No	No Yes	Yes No	Yes Yes	Other Other	2021	2022
At least one board member oversees CS risks	199	69	101	223	82	46.3%	51.5%
Staff member responsible for CS directly reports to board	141	95	100	310	28	47.8%	62.0%
Any of board received training	190	41	76	189	178	39.3%	44.8%
CS included in most recent annual statement	233	34	45	48	314	15.9%	19.7%

(b) **Board CS Discussion and Business Integration**

Change from 2021	Decrease	Same	Increase	N/A	2021 Total	2021	2022
Frequency board discussed/updated on organisation's cyber security	151	223	180	120	674		
Never	–	49	51	7	107	15.9%	13.8%
Once a year	23	37	63	6	129	19.1%	17.4%
Once every 6 months	29	23	41	4	97	14.4%	16.2%
Quarterly	50	73	23	15	161	23.9%	26.9%
Monthly	42	40	2	6	90	13.4%	14.7%
Weekly or more	7	1	–	1	9	1.3%	1.2%
N/A				81	81	12.0%	9.9%
Board integrates cyber risk considerations into wider business areas	142	192	129	211	674		
Strongly disagree	–	0	14	5	19	2.8%	1.8%
Tend to disagree	3	12	31	16	62	9.2%	9.1%
Neither agree nor disagree	13	29	47	17	106	15.7%	18.5%
Tend to agree	60	95	37	15	207	30.7%	32.8%
Strongly agree	66	56	–	7	129	19.1%	19.3%
N/A				151	151	22.4%	18.5%

Table C.3 provides results showing the relationship between a board's involvement in cybersecurity and incident frequency. The variables in the table capture aspects of board governance, knowledge, and interaction with cybersecurity. I do not include the control for how frequently the board discusses cybersecurity from Table C.2, as this variable is not significant and including it sharply reduces the sample.

Table C.3: Board Involvement Measures and Incident Frequency

	Dependent variable: incident								
	Including phishing			Excluding phishing			Only phishing		
	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6	0 vs 1-6	0-1 vs 2-6	0-2 vs 3-6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\geq$ 1 board member role includes CS oversight	0.050	0.105	-0.075	0.020	0.103	-0.060	0.084	0.184	0.018
	(0.052)	(0.077)	(0.096)	(0.073)	(0.105)	(0.101)	(0.071)	(0.132)	(0.140)
Staff member responsible CS reports directly to board	-0.018	0.093	0.070	-0.094	0.119	0.018	-0.048	-0.358***	-0.280*
	(0.050)	(0.086)	(0.116)	(0.078)	(0.108)	(0.116)	(0.075)	(0.131)	(0.153)
Any board members received training	0.026	-0.097	-0.160	0.027	-0.110	-0.053	0.014	-0.092	-0.160
	(0.043)	(0.079)	(0.109)	(0.081)	(0.105)	(0.131)	(0.069)	(0.147)	(0.150)
CS in most recent annual report	-0.039	0.024	0.085	0.003	0.052	0.020	-0.039	0.024	0.245**
	(0.049)	(0.067)	(0.094)	(0.081)	(0.083)	(0.114)	(0.074)	(0.145)	(0.109)
Board integrates CS considerations (rel to "Strongly agree")									
Agree	0.020	0.087	0.087	-0.116	0.011	0.043	-0.020	0.071	0.087
	(0.060)	(0.076)	(0.088)	(0.085)	(0.099)	(0.099)	(0.079)	(0.151)	(0.131)
Neither agree/disagree	0.101	0.142	0.018	-0.003	0.084	0.117	-0.074	-0.052	-0.199
	(0.080)	(0.105)	(0.119)	(0.122)	(0.136)	(0.131)	(0.094)	(0.164)	(0.152)
Disagree	0.085	0.083	0.233*	0.039	0.007	0.161	0.027	0.071	0.005
	(0.066)	(0.117)	(0.126)	(0.105)	(0.134)	(0.152)	(0.094)	(0.184)	(0.167)
Strongly disagree	0.047	0.187	0.199	-0.205*	0.071	0.429*	0.224	0.277	-0.191
	(0.072)	(0.204)	(0.203)	(0.121)	(0.297)	(0.236)	(0.152)	(0.335)	(0.330)
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	571	571	571	505	505	505	636	213	213
Adjusted R^2	0.486	0.286	0.277	0.567	0.383	0.272	0.351	0.251	0.121

*p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the organization level. The significance level, direction, and magnitude of the coefficients are robust to including type (i.e., business or charity), sector, or size fixed effects.

Having a board member with oversight of cybersecurity does not have a significant relationship or general directional pattern with the experienced frequency of incidents. However, having a staff member responsible for cybersecurity directly report to the board appears to have a moderately strong negative relationship with the frequency of incidents. It is not immediately obvious why this relationship exists. One possibility is that having someone from the IT team report to the board may indicate an organization that is more conscious of cybersecurity. In turn, this is likely to foster an organizational culture aware of cyber risks, leading to more skepticism and less phishing incidents. Regardless, the large effects found in Regressions 8 and 9 could simply be a product of the minimal data on phishing incident frequency.

Though insignificant across all regressions, it is worth pointing out that having cybersecurity

training for board members seems to have a negative relationship with the likelihood of high frequencies of incidents (Regressions 2-3, 5-6, and 8-9). This could suggest that requiring cybersecurity training for an organization's board is helpful in reducing the number of cyber incidents. This does not seem to be the case for entirely preventing incidents. However, reducing the overall number of incidents is certainly a worthwhile goal for an organization. More data needs to be collected and assessed to determine whether board cybersecurity training is helpful.

Including cybersecurity in an annual report is generally unrelated to the experienced frequency of incidents. However, Regression 9 shows a strong positive relationship between doing so and the likelihood of experiencing monthly or more phishing incidents. It is worth noting the small sample size supporting this result, though it is fairly intuitive: organizations with valuable digital assets are likely to experience a high degree of phishing, and it is likely an obligation to include something regarding cybersecurity in an annual report to give stakeholders a sense of the risks faced by the organization.

Generally, it seems that more disagreement that an organization's board integrates cyber concerns across business areas is associated with a higher likelihood and frequency of incident. The exception to this is in Regression 4, where the large negative coefficient on "Strongly disagree" is weakly significant relative to the reference group of "Strongly agree." This is likely due to a smaller number of organizations reporting "Strong" assessments, as the coefficients on "Agree" and "Disagree" are consistent with the rest of Table C.3 where lower incident frequencies are generally associated with more agreement that the board integrates cybersecurity considerations into wider business concerns.

C.3 Cybersecurity Terms

Table C.4: Glossary of Cybersecurity Terms

Term	Definition
Business Continuity Plan (BCP)	"A Business Continuity Plan is the plan for emergency response, backup operations, and post-disaster recovery steps that will ensure the availability of critical resources and facilitate the continuity of operations in an emergency situation." (SANS, 2024)
Cloud computing	"A model for enabling on-demand network access to a shared pool of configurable computing capabilities or resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." (NICCS, 2024)
Data breach	"The unauthorized movement or disclosure of sensitive information to a party, usually outside the organization, that is not authorized to have or see the information." (NICCS, 2024)
Denial of Service	"A cybercrime in which the attacker floods a target with internet traffic to prevent users from accessing connected online services and sites." (NICCS, 2024)
Impersonization	"An attack type targeted phishing attack where a malicious actor pretends to be someone else or other entities to steal sensitive data." (NICCS, 2024)
Incident	"An occurrence that actually or potentially results in adverse consequences to (adverse effects on) (poses a threat to) an information system or the information that the system processes, stores, or transmits and that may require a response action to mitigate the consequences." (NICCS, 2024)
Incident Response Plan (IRP)	"A set of predetermined and documented procedures to detect and respond to a cyber incident." (NICCS, 2024)
Malware	"Software that compromises the operation of a system by performing an unauthorized function or process." (NICCS, 2024)
Phishing	"A digital form of social engineering to deceive individuals into providing sensitive information." (NICCS, 2024)
Ransomware	"A malware designed to deny a user or organization access to files on their computer." (NICCS, 2024)
Risk Register	"A repository of risk information including the data understood about risks over time." (NIST, 2024)
Threat Intelligence	"Threat information that has been aggregated, transformed, analyzed, interpreted, or enriched to provide the necessary context for decision-making processes." (NIST, 2024)
Virtual Private Network (VPN)	"A restricted-use, logical (i.e., artificial or simulated) computer network that is constructed from the system resources of a relatively public, physical (i.e., real) network (such as the Internet), often by using encryption (located at hosts or gateways), and often by tunneling links of the virtual network across the real network." (SANS, 2024)
Virus	"A computer program that can replicate itself, infect a computer without permission or knowledge of the user, and then spread or propagate to another computer." (NICCS, 2024)