SRI International

SRI International Work on Cybereconomic Incentives for the Department of Homeland Security Science and Technology Directorate Cyber Security Division

January 31, 2015

Prepared for:
Dr. Joseph Kielman
Cybersecurity Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Bincy Ninan-Moses, Roland Stephen, Ph.D., Lucien Randazzese, Ph.D., and Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development
David Balenson, Ulf Lindqvist, Ph.D., and Zachary Tudor
Computer Science Laboratory

Primary Contact:
Lucien Randazzese, Ph.D.
SRI International
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
lucien.randazzese@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

1 Overview

In the nearly year and a half since the President issued Executive Order (EO) 13636 on *Improving Critical Infrastructure Cybersecurity*[1] and Presidential Policy Directive (PPD) 21 on *Critical Infrastructure Security and Resilience,*[2] there has been a great deal of policy discussion and analysis of the incentives associated with cybereconomics. Much of this assessment has focused on how incentives might influence adoption of the voluntary framework for reducing cyber risks to critical infrastructure developed by the National Institute of Standards and Technology (NIST). As part of this focus on incentives, the Departments of Homeland Security (DHS), Commerce, and Treasury identified potential incentives for infrastructure owners and operators to adopt the NIST framework.

The initial analysis by the executive branch frames incentives in terms of marginal economic costs and benefits. SRI International provided input to the DHS Science & Technology (S&T) Directorate's Cybersecurity Division (CSD) as CSD set out to define a long-term research program around the topic of cybereconomic incentives (CEI). In considering the strategic direction of such a research program, SRI proposed taking a broader perspective on the subject of cybereconomic incentives than had been followed to date. Specifically, SRI advocated for a view of incentives that explicitly considers behavioral factors that affect human decision making in the context of cybersecurity, and proposed a set of related activities aimed at bootstrapping a broader, long-term research enterprise focused on these behavioral factors.

The proposed activities included reviews of current cybereconomic incentives research and policy-focused behavioral science research, used to inform a proposed research agenda in CEI, as well as development of a field experiment aimed at demonstrating the utility of the behavioral approach in understanding cybereconomic decisions. In total SRI produced a set of five analyses and documents, collected here in a single source.

The following documents were produced by SRI for DHS CSD and are included in this compendium:

1. **Concept Paper: Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives** – A concept paper that outlines a framework for research in cybereconomic incentives that launches from standard microeconomic analysis into new opportunities for research emphasizing behavioral sciences.

[1] The White House, *Improving Critical Infrastructure Cybersecurity*, Executive Order 13636, February 12, 2013.

[2] The White House, *Critical Infrastructure Security and Resilience*, Presidential Policy Directive 21, February 12, 2013.

2. **Literature Review: Current Research in Cybereconomics** – A review of the current research in cybereconomics. This review is the first of two research reviews in this work stream.
3. **Literature Review: The Application of Behavioral Research in Public Policy** – A review of the applications of behavioral science research in policy and management areas outside of cybersecurity.
4. **Proposed Research Agenda for Cybereconomic Incentives** – A proposed research agenda for the field of cybereconomic incentives, focusing on both the near-term and long-range research needs of DHS's mission of enhancing the security and resilience of the nation's critical information infrastructure.
5. **Proposed Research Experiment for Cybereconomic Incentives** – A proposed research experiment intended to evaluate how small and medium businesses (SMBs) involved with the nation's critical infrastructure respond to incentives to improve their cybersecurity, including incentives with strong behavioral components.

2 Summary of Collected Work

Two themes underlie the collected work presented. The first stresses the fact that study of decisions made regarding cybersecurity, by individuals and organizations, cannot be assessed in terms of responses to economic incentives, or perceived costs and benefits alone. Research must also consider the behavioral aspects of human decision making. While a growing share of the existing cybereconomic incentives research has started to focus on the behavioral aspects of cybersecurity, in almost all cases it considers personal behavior around privacy, and fails to address how people behave when they serve as the stewards of an organization's collected data or infrastructure.

Second, our work represents a call for research that can be of more direct benefit in informing public policy and management decisions regarding cybersecurity. In large part this means a more empirical approach to evaluating what actually people do and how they process information when facing decisions regarding how to behave and how to invest in cybersecurity. For example, while much careful work in microeconomics highlights the public good nature of investments in cybersecurity, and concludes that firms left to rely on their own private return on investment (ROI) for investment decisions will underinvest in cybersecurity, little is known about how firms actually go about making financial decisions to invest (or not to invest) in security. Nor is much known about what external factors, including behavioral factors, may influence these decisions. Absent this understanding, crafting effective policy becomes much more difficult.

Section 2.1 below outlines the arguments made in SRI's original concept paper on the theme of cybereconomic incentives. Section 2.2 represents the first of the two literature reviews completed, and reviews, summarizes, and discusses current CEI research. Section 2.3, representing the second literature review, examines the applications of behavioral research in policy areas outside of cybersecurity. We have integrated the insights from these two streams of

research to describe the joint role cost-benefit analyses and behavior influences have on decision making, and the resulting implications for cybersecurity policy. This integrated policy approach is described in detail in Section 2.4. The final two sections of this summary, 2.5 and 2.6, describe SRI's proposed CEI research agenda and the potential CEI field experiment. Copies of the complete reports for the work summarized in each of these sections follow this summary document.[3]

2.1 Concept Paper: Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives

Experience shows that bad actors are able to circumvent technical approaches to cybersecurity through behavioral means, or "social engineering." It is reasonable then to conclude that cybersecurity can be bolstered more effectively through means that address behavioral considerations, not simply technology. Such behavioral approaches would rely not on manipulation and misdirection, as in the case of most social engineering exploits, but rather by changing the decision-making environment of system developers, vendors, service providers, and end-users.

While standard microeconomics represents a useful starting point in understanding this decision-making environment – people do often calculate, or attempt to calculate, the gains and losses associated with their choices – in practice there are limits to anybody's ability to reason about the world. Examples abound in real life of this "bounded rationality" by which people seem to make choices without perfect (or sometimes any) regard for the costs they will incur or the payoffs they will receive. A holistic approach to researching cybereconomic incentives needs to go beyond the construction of economic cost-benefit models, and leverage knowledge gained from other social science research on incentives and behavior.

Behavioral influences on decision making include a range of cognitive considerations outside of strict cost-benefit accounting, and include, *inter alia*:

- The effects of context
- Bias towards exiting beliefs
- The desire for fairness
- Peer behavior and influence
- Asymmetrical perception of losses versus gains

The SRI concept paper also outlines a three-component stakeholder map for use as a framework for understanding how cybereconomic incentives impact the range of entities

[3] Excluding Section 2.4 Integrated Policy Approach, which was developed after the five original project documents were competed.

involved in deploying, using, and defending critical infrastructure. The components of this framework are:

- Major players, including network operators, Fortune 500 enterprises, federal agencies, etc.
- Entities further down the value chain, including local service providers and partners, tier two/tier three supply chain enterprises, state and local governments, etc.
- Consumers and employees (often one and the same).

The concept paper concludes with an outline for a research and planning process with two thrusts – one in formulating a research agenda in cybereconomics, and one in designing a field experiment ("proof-of-principle") in the application of behavioral economics to the design of incentives around cybersecurity practices. As inputs to these tasks, the research agenda also sets out to conduct the two literature reviews, reviewed next.

2.2 Literature Review: Current Research in Cybereconomics

The first of the two reviews surveys the existing research in cybereconomics. The initial work in cybereconomics, which emerged in the early 2000's, was in the field of microeconomics, and featured theoretical mathematical models that highlighted the public good nature of investments in cybersecurity. Given the externalities inherent in protection systems, this work concludes that firms left to rely on their own private ROI for investment decisions will underinvest in cybersecurity. While this research provides occasional insight into the economic dynamics of security investment, it is often theoretical in nature, producing results that are either common sense in nature or vague in their implications for policy.

This microeconomic work also largely ignores differences between major players (such as large corporates and federal government agencies) who tend to invest considerable resources in cybersecurity and secondary players in the value chain who invest much less and are thus more vulnerable to attack. More generally, there is little consideration of just how far from optimal investment is overall, as a whole or on the part of individual entities. The policy implications of a nation that is grossly underinvesting in security are obviously different than those in the case where investment is close to optimal.

In reaction to the limitations of the theoretical research, research focusing on actual behavior – individual and organizational – has become a central focus in studies of cybereconomics in the last several years. This research includes some theoretical work, but also makes novel use of specific datasets (e.g. consumer website usage) to examine behavior empirically, as well as formal experiments on how people behave in various cybersecurity settings. Some of the most interesting cybereconomic work is taking place in this domain. For example, two sets of researchers have shown that people are less likely to behave offensively online when their actual identities are shown along with their comments. Similarly, a number of researchers have

shown that people are more likely to divulge sensitive information when they are told other people have, or when they feel they are in control of how information is revealed, *even when that control has no impact on who ultimately can gain access to the information being divulged*. The implications of such findings for cybersecurity would seem clear.

Real world data has also been used to investigate the motives and activities of cybercriminals. Data from captured infrastructure used in the conduct of actual cybercrime shows that cybercriminals seek to keep scams going as long as possible, even to the point of issuing refunds to complaining (i.e. scammed) customers in effort to avoid detection. This type of research represents a novel and potentially useful approach in analyzing cybercriminal motivations and behavior.

While behavioral research has made some headway into our understanding of the decisions people make with respect to sensitive information, most of it focuses on personal privacy issues, and less on how people behave when they serve as the stewards of someone else's sensitive information, for example as employees of companies or government organizations with knowledge of proprietary or classified data, or in control of critical cyber infrastructure. Employees, of course, represent a significant source of cyber risk, and so looking specifically at their behavior, and why they do or do not comply with security policies, would make a fruitful path of research.

There is some preliminary research, reported on in this review, of the drivers of good security practice at the organizational level, but less so at the individual level. One interesting result on organizational behavior, one that shows organizations to be as subject to psychological biases as are people, is that organizations will try to reduce the amount of spam they are responsible for when their spam levels are publically reported, but also that they put less effort into this reduction when other groups are reported as considerably worse offenders.

The cybereconomics research review also highlights the lack of data as a critical methodological handicap to cybereconomic research, one often lamented by researchers. Among the areas for which there is a need for more data are: cybersecurity policies and activities; the costs associated with cybersecurity; and security incidents and their outcomes, both technical and non-technical. The disincentives for sharing data are well known, but the common benefits to more widely available information are also clear.

2.3 Literature Review: Applications of Behavioral Research in Public Policy

The second SRI literature review evaluated the application of the insights from behavioral science to a wide range of policy and management settings outside of cybersecurity. Each of the behavior research applications considered in this review involve circumstances, policy challenges, or policy goals with parallels to the circumstances, challenges, and goals associated with improving cybersecurity, particularly the incentives associated with cybersecurity. Thus

such research is likely to provide insight into improving the security and resilience of the nation's critical infrastructure.

Some of these real world applications of the insights of behavioral science described are quite novel, but many are actually very straight forward – some may even appear obvious. What makes them noteworthy is their level of effectiveness in achieving the behavior-changing goals they set out to address. The absolute number of examples is large and growing rapidly. The examples presented in the review do not compose a comprehensive list. Rather, they are representative illustrations of how behavioral science insights are being applied to real world problems, and many come from governments and organizations at the forefront of exploiting behavioral research. Application areas covered in the review include public health, crime prevention, financial decision making, consumer marketing, energy efficiency, college admissions, tax collection, and many others.

In addition to the many examples it provides, SRI's second literature review describes the development of behavioral sciences as a field, and highlights its recent ascendency in influencing several governments in a broad range of policy issues. The scholars that contribute to this area of research tend to be psychologists, sociologists, neuroscientists, and other scientists outside the field of economics per se. Nevertheless, the field of science that deals with how cognitive, social, and emotional factors affect human decision making is often referred to as behavioral economics. We use that term in the review, but also refer more generally to behavioral sciences and behavior research when discussing the insights from this research and their applications to policy.

The review also discusses the DHS Integrated Task Force view of incentives in regard to adoption of the NIST cybersecurity framework, the assumptions made about the nature of decision making reflected in this traditional microeconomic view, and the potential limitations to this view. We then go on to describe how a broader behavioral view of incentives differs from this traditional view.

The behavioral science applications are organized into six broad categories. The six categories were chosen as an effective way to organize what is a large number of specific insights into human cognition, many of which are related to one another. A number of similar ideas, for example such as anchoring and priming, are sufficiently similar that they have been grouped together. Table 1 below summarizes each of the six categories of behavioral insight, and describes their implications for cybersecurity incentives.

Decision Factor	Insight from Behavioral Science	Implications for Cybersecurity Incentives
Choice	The process of choosing is difficult, and people will often make choices in a way that minimizes the effort in making the choice, with little or no consideration of the actual options	Superior security options may not be selected if that selection required too much analytical effort
Loss	Humans are risk averse when faced with a loss and risk accepting when faced with a gain when each are of equal value	People may react more strongly to incentives conveyed as protection from losses than those seen as rewards or payments
Anchoring	Our evaluations and behaviors are more affected by recent information, experience, or stimuli	Measuring and broadly promulgating the nature and consequences of security breaches could help overcome anchoring -based complacency with respect to security
Representativeness	People draw incorrect conclusions about causation and distribution when evaluating random data	Assessment of the location (in time and space) of cybersecurity risk and the impact of this risk may be biased
Control	People will do things they "do not want to do"	Policymakers and organizations should not assume people will avoid behavior merely because a behavior or its consequences are detrimental to people's self-interest
Peer Influence	People are susceptible to peer pressure and rely on peers as sources of low-cost information about how to choose	Cybersecurity-related incentives that require group-wide compliance or performance may be more effective than incentives aimed at individuals alone

Table 1: Implications of Behavioral Science for Cybersecurity

The concepts summarized in the table above and in this review generally portray a wide range of situations in which human decision making is influenced by psychological rather than purely objective analytical factors. As the review illustrates, policymakers are exploiting these behavioral biases to influence behavior in many policy areas. Equally wide ranging has been the geographic scope of the policy application of behavioral theory; its concepts have been used by governments in North America, Europe, and Asia. Obviously, there are lessons for managing and influencing the incentives that govern cybersecurity behavior and investment.

Some of the cybersecurity applications of the ideas presented in this review may be quite simple – for example, the lessons of choice architecture, choice fatigue, and the power of defaults suggests design principles that make the highest level of security a universal default. Other applications might be more difficult in the real world to apply. For example, while shaming has proved effective in getting people to pay their taxes, in the cybersecurity context it raises questions about potential target identification for would-be attackers.

2.4 Integrated Approach for CEI Policy Research

Over the long run, developing effective cybereconomic incentives policies will require combining lessons from traditional economics, behavioral research, and deep technical subject matter expertise in cybersecurity. In this section we consider the first two of these requirements: insights from economics and those from behavioral research. The discussion in this section reflects work done after the five project documents described in the overview, and so is not included anywhere else in this compendium other than this section.

An integrated approach for CEI policy research begins with the premise that cost-benefit analyses, also referred to as microeconomic analyses or simply economic analyses, play a central role in cybersecurity related decision making. The quality of purely economic decisions, however, is greatly affected by the availability and quality of relevant information. Decision quality can also be impacted by the presence of behavior influences that may jeopardize the soundness of cost-benefit decisions even when good cost-benefit information is available. These concepts are illustrated in Figure 1 below.

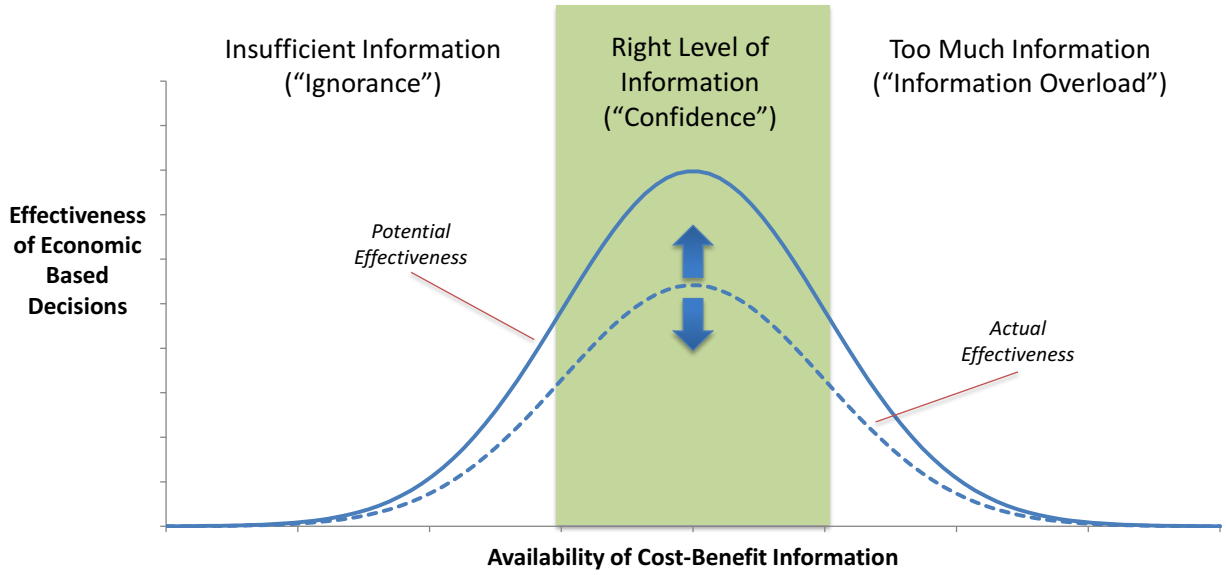


Figure 1: Quality of Decisions Made via Cost-Benefit Analysis

The x-axis of Figure 1 represents the availability of cost-benefit information relevant to a specific decision. The y-axis measures the quality or effectiveness of decisions that can be made given the availability of information. On the left hand side of the figure, insufficient information is available, and decision makers are forced to operate in a regime of ignorance. In such situations the effectiveness of decisions is likely to be low.

In the real world, especially in situations involving complex interactions between people, technology, organizations, and the marketplace, information overload is often a problem. In

situations of information overload, information is so abundant, and often so contradictory, diffuse, and rapidly changing, that its effective use in decision making becomes impossible. As a practical matter, information overload has the same impact as insufficient information; both situations are characterized by poor decision making. Only when an amount of information is available that allows for accurate evaluations of the costs and benefits of a decision – the center region of Figure 1 – can an effective cost-benefit decision be made with confidence.

Even in cases in which there is an appropriate amount of information available, however, decisions can still be influenced, and potentially jeopardized, by behavioral influences. Consider the example of password sharing. An employee, absent any peer influences, will likely conclude that sharing his password bears a high potential cost (severe employer or even legal sanctions if his password sharing is discovered) relative to whatever benefit may accrue from sharing his password with a colleague (short-lived appreciation of the borrower). In such a situation, the employee would decide to not share his password based solely on analysis of the costs and benefits of sharing.

In real life we know that password sharing is common. It is so, in part, because some prospective borrowers are persistent. They may claim a real or contrived emergency need for a password, and convince, via peer influence, their colleagues to ignore their internal cost-benefit analysis. The ability of behavioral factors to degrade otherwise good cost-benefit base decisions is represented by the lower dashed line in Figure 1. While there is potential to make good decisions when good cost-benefit information is available, there is no guarantee that good decisions are made.

Figure 2 highlights the role of policy in the decision dynamics of the different regimes of information availability depicted in Figure 1. When good cost-benefit information is scarce, the role of policy is to stimulate information creation and dissemination, or possibly to create and provide missing information directly, thus moving the decision maker from ignorance to confidence. Conversely, when the information environment is crowded and confusing, the role of policy is to simplify the information environment, for example through the creation and/or promulgation of standards. Here again the effect is to move decision makers towards confidence, where they can make more effective cost-benefit decisions.

When available information allows decision makers to make effective economic based decisions, the role of policy is to influence incentives already in place, in a manner that increases the likelihood of meeting policy objectives. Consider, for example, the provision of tax rebates that lower the effective cost of investment in cybersecurity. Introducing such a policy will have the most beneficial, and predictable, outcomes when there is already sufficient information for entities to make sound economic decisions.

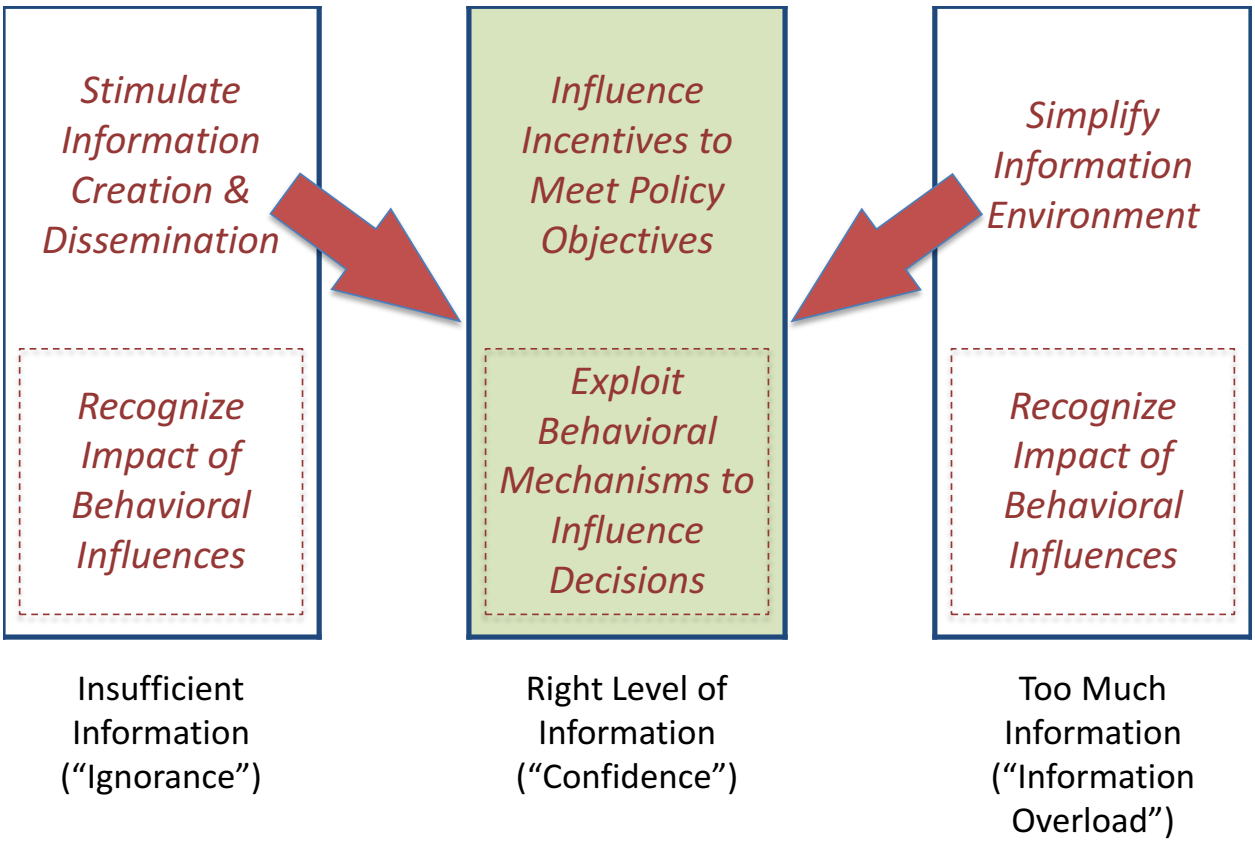


Figure 2: Role of Policy in Cybereconomic Incentives

In addition to the capacity of policy to affect economic incentives, there is also a role for policy in recognizing and exploiting the power of behavioral mechanisms to change behavior. This is true in cases in which the information environment is one of ignorance or overload, but also when decision makers have the information needed to make good decisions. As the password sharing scenario indicates, behavioral influences can play a role even when the information environment is conducive to good decision making. In these situations, behavioral influences on human cognition and choice can reinforce good decisions, or jeopardize them. It is the role of policy to exploit what is known about behavioral influences to ensure that they are used to improve the likelihood of good decision making. The dashed red boxes of Figure 2 highlight the policy opportunities made possible by what we know of behavioral influences on human cognition.

Practical constraints on economic oriented policies constitute another reason there is a role for behavioral based policy even in environments which lend themselves to good microeconomic based decision making. It is often financially unrealistic or otherwise impractical to implement economic based incentive policies needed to encourage specific decisions. Consider again the example of password sharing. The government could, say, in theory, require a mandatory ten-year prison sentence for anyone sharing a commercial password. Such a law would dramatically affect people's economic incentives regarding password sharing, and in response

to the new incentives created by such a law, the frequency of password sharing would decline dramatically. But such a policy is obviously impractical, creating opportunities for alternative ways to influence behavior.

The key to effective policy creation in the long run is to recognize that behavior, at both the individual and organizational level, can be influenced. Natural economic incentives already at play can be exploited, pushing the economic costs and benefits of decisions in the direction desired by policy. And beyond the manipulation of economic incentives, there is a role for policy in affecting decisions by exploiting the growing body of knowledge of the behavioral influences on decision making.

2.5 Proposed Research Agenda for Cybereconomic Incentives

The research agenda outlines a proposed cybereconomic incentives (CEI) research agenda to guide the DHS CSD in identifying and supporting areas of CEI research that will advance CSD's mission of enhancing the security and resilience of the nation's critical information infrastructure. The proposed research areas are intended to identify a further body of research that builds on and significantly extends the current research, and are organized into six broad research categories

- Category 1: The Economics of Cybersecurity Investment Incentives
- Category 2: Individual Incentives & Behavior
- Category 3: Organizational Incentives & Behavior
- Category 4: Attacker Incentives & Behavior
- Category 5: Cyberinsurance and Cyber Liability
- Category 6: Cybereconomic Incentives Data Collection

There are interrelationships among the categories, with some inevitable overlap in coverage. For example, the study of individual behavior (Category 2) might inform certain questions about how organizations behave (Category 3). Category 6 is cross-cutting, and has implications for research in each of the other five categories.

Within each research category, broad areas of research are described and a number of specific exemplar research questions and topics are identified. A total of ten specific research areas are described across the six proposed categories of research. For some research areas, potential frameworks, databases, or other tools whose development would be useful to policy makers are also identified. The last section of this document proposes a few topics for consideration as potential early priorities of the research agenda.

Parts of the proposed research agenda identify issues that are addressable in the short term and have a fairly applied orientation. Other areas pose questions that require more fundamental evaluations of human and organizational behavior, and are therefore generally longer-term in

focus. Accordingly, in describing the research areas below, the specific questions associated with each are organized according to their short- versus long-term focus. This distinction is at best a general guide, as there may be shorter and longer term paths of research for many of the individual questions.

The proposed research areas address the cybereconomic incentives affecting all three of the stakeholder groups described in the concept paper's analysis framework: major players; smaller entities further down the value chain; and consumers and employees. The six categories of proposed research also align closely to the fields of research described in the cybereconomics literature review, though the correspondence is not exact. In some cases research that was topically similar but methodologically distinct was separated into distinct sections of the literature review. This proposed research agenda is largely agnostic on issues of research methodology, and accordingly aggregates research streams focusing on the same topic.

The final section of the proposed research agenda section identifies a small number of specific topics for consideration as potential early priorities. Three criteria were used to identify these potential priorities:

- The relative ease with which short-term progress can be made for the proposed topic
- The amount of insight any quick-to-emerge findings would provide policymakers and other researchers
- The potential of these topics to spur interest and research activity in new areas of cybereconomic incentives

Based on these criteria, the following initial priorities were proposed. All of the priority topics but one fall in the shorter-term question category; the one exception focuses on framework and tool development.

From **Research Area #1** (Improved understanding of current patterns of investment in cybersecurity):

- What are actual levels of investment within the overall system, and how is total spend composed?
- How does spend vary by cybersecurity activity within the Identify-Protect-Detect-Respond-Recover cycle?
- How do organizations evaluate the return on investment (ROI) on cybersecurity, and what ROIs are being realized in practice?
- Frameworks and tools for comprehensive quantification of the costs of cybersecurity and cost of data breaches

From **Research Area #2** (The impact of market forces on cybersecurity investment and behavior at firms):

- How do customers react to security breaches?
- How do customer reactions to breaches, and the extent to which they are negatively impacted by such breaches, affect firm security investment and behavior?

From **Research Area #3** (Behavioral mechanisms affecting individuals when they are responsible for the data and data assets of others):

- What evaluations can be done of existing efforts to promulgate appropriate cybersecurity behavior (e.g., Stop.Think.Connect.), and how do these inform understanding of how behavior can be changed through programmatic approaches?

From **Research Area #5** (The role of trust in cybersecurity):

- How do organizations evaluate the trustworthiness of individuals with access to information technology and industrial control systems, and where and why do these evaluations fail?

From **Research Area #7** (Cybercriminal behavior & incentives):

- Are there non-financial incentives that will motivate so-called white hat hackers to improve cybersecurity beyond participation in open-source bounties?

From **Research Area #10** (Cybereconomic Incentives Data Collection):

- What lessons are there from other policy domains, e.g. the Sarbanes–Oxley Act and financial disclosure?

2.6 Proposed Cybereconomics Research Experiment

The SRI work on cybereconomic incentives culminated in the development of a proposed experiment that directly addresses behavioral influences on corporate decisions regarding cybersecurity investment. To our knowledge, the proposed experiment is the first such experiment focused on understanding the behavioral influences on people in real-world organizations making real-world decisions about cybersecurity.

The experiment presented here is still at the notional stage, and subject to change as more work is done to design and ultimately conduct it. The proposed experiment evaluates how small and medium businesses (SMBs) respond to an offer of a no-cost assessment of their potential cybersecurity vulnerabilities. The eventual experiment may or may not focus on SMB responses to a cybersecurity assessment. Essential to the experiment's final design is its examination of actual corporate decisions about cybersecurity and the inclusion of behavioral factors in its treatment groups.

The current proposal considers as subjects SMBs involved with the nation's critical infrastructure. Subjects will receive a cybersecurity assessment offer (to be conducted by an appropriate external organization) that incorporates incentives designed to improve subject interest in going through with the assessment, including incentives inspired by behavioral

science. A control group will receive the assessment offer with no incentives. Analysis of response rates will highlight the relative effect of the incentives tested.

The experiment is conceived of as comprising two stages:

1. Pilot stage conducted with a narrow set of SMB respondents to evaluate response rates and degree of insight coming from pilot responses
2. Full rollout with a broader set of SMB respondents, during which treatment effect differences across various firmographic variables, such as firm size and industry, are evaluated

The SMB subjects of the proposed experiment will be divided into different treatment groups, each of which will be offered the same cybersecurity assessment. Subjects in the control group will receive the assessment offer without any additional incentive. Subjects in each of the treatment groups will receive the assessment offer plus an additional incentive to agree to the assessment. One incentive is economic in nature, while the remaining will key on potential behavioral influence of decision making: peer influence, anchoring, and loss aversion. These behavioral factors were selected because they lend themselves to evaluation in the type of experiment proposed, i.e. one in which an offer is made, subject to some perceived cost and benefit.

The assessment itself will be developed within the context of the voluntary framework for reducing cyber risks to critical infrastructure, developed by the National Institute of Standards and Technology (NIST). In other words, it will assess how knowledgeable of and compliant with the NIST framework participants currently are. While we hope the proposed experiment will have the effect of encouraging adoption of the NIST Framework; ultimately its objective is to attract greater interest in research of this kind by demonstrating that exploiting behavioral science can materially affect how individuals within firms make cybersecurity decisions.

SRI International

Concept Paper:
Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives

December 18, 2013

Prepared for:
Dr. Joe Kielman
Cyber Security Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Roland Stephen, Ph.D. & Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development
SRI International
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
roland.stephen@sri.com
jeffrey.alexander@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

Background: Policy Developments in Cybereconomics

Cybereconomics as a topic on the U.S. federal R&D agenda for cybersecurity has emerged fairly recently. In the April 2006 *Federal Plan for Cyber Security and Information Assurance Research and Development*, produced by the CSIA IWG, the topic of economics was hardly mentioned. The 2009 National Cyber Leap Year exercise brought the field of cybereconomics to prominence. The "arms race" between bad actors and cybersecurity tool developers is a losing game for those parties attempting to ensure security and assurance. As stated in several public meetings, the traditional approach to promoting cybersecurity is fundamentally broken.[1]

Experience shows that bad actors are able to circumvent technical approaches to cybersecurity through behavioral means ("social engineering"). It is reasonable that cybersecurity could be bolstered by means which address behavioral considerations, not simply technology. These behavioral approaches would rely not on manipulation and misdirection, as in the case of most social engineering exploits, but rather by changing the decision-making environment of system developers, vendors, service providers, and end-users. In other words, economic analysis could assist in designing policy interventions, or incentives, to encourage better choices related to cybersecurity.

The 2010 "Cybersecurity Game-Change R&D Recommendations" from the CSIA IWG outlined the parameters of a research agenda for cybereconomics. That document, following on the National Cyber Leap Year co-chairs' report, frames the topic in terms of "misaligned incentives and misallocated resources." In that context, research on cybereconomics is driven primarily by classical economic analysis. Assuming that actors in the system (including vendors, system users, and bad actors) are rational, and that information flows are both efficient and comprehensive, it should be possible to design a set of interventions that shifts the costs and benefits of cybersecurity measures so that all participants see the value in creating a more secure environment. In other words, incentives can be rationalized in a way that ensures that "crime does not pay."

This agenda carried through to the 2011 *Strategic Plan for the Federal Cybersecurity Research and Development Program*. The Cyber Economic Incentives component of that plan argues that to achieve a secure cyberspace, "the projected benefits must be quantified to demonstrate that they outweigh the costs incurred by the implementation of improved cybersecurity measures." The research agenda is structured around the research needed to provide a clear business case for secure behaviors—validated data on the costs of cybersecurity measures, clear metrics to show the advantage gained by adopting good cybersecurity measures, and a

[1] See, for example, the panel organized by the Washington, DC chapter of The Internet Society titled "Cybersecurity 2020: Is There a Better Way to Protect the Internet?," held on May 23, 2012 at SRI's offices in Arlington, Virginia.

resulting model which will indicate areas where policy intervention is required to make that cost-benefit calculation clear and compelling. Assuming that this research effort is successful, it should be possible to design a framework describing proper cybersecurity measures and activities, and institute policies which will convince target audiennces to adopt that framework. This is the goal described in a recent blog post by Michael Daniel, Special Assistant to the President for Cybersecurity[2].

The standard microeconomic approach is a useful and powerful point of departure when thinking about behavior. People often calculate the gains and losses that may arise from one choice or another. Using payoffs as an incentive to make them act in a certain way can often shape behavior. In practice, however, there are limits to anybody's ability to reason about the world (so-called "bounded rationality"). As a result, there are kinds of real-world examples of people who seem to choose without perfect regard for the payoffs they will receive. Behavioral economics is designed to describe and, where possible, explain such actions that are systematically different from the standard, rational approach.

The research perspective of behavioral economics attempts to develop models which account for the irrational aspects of human decision-making, drawing from fields such as cognitive psychology and political science. A more holistic approach to researching cyber economic incentives would go beyond the construction of straightforward cost-benefit models, and leverage knowledge gained from other social science research on incentives and behavior to find opportunities for new types of financial and non-financial incentives. This is consistent with work being done in other policy domains. For example, a recent workshop convened by the National Institute of Aging, in collaboration with the White House Council of Economic Advisers (CEA), the White House Office of Science and Technology Policy (OSTP), and the Association for Psychological Science (APS), explored ways in which psychological science and behavioral economics can inform future policy decisions designed to encourage beneficial behavioral change in a population.[3]

This concept paper outlines an alternative framework for research in cyber economic incentives which launches from standard microeconomic analysis into new opportunities for constructing field experiments using behavioral economics and social science to design new types of cyber economic incentives. We present below a framework for analyzing the actors in the system whose behavior in cybersecurity will be the focus of such experiments. We then discuss lessons learned from previous work on behavioral economics in other policy domains, and construct a plan for a potential "proof-of-principle" experiment in cybersecurity behavior.

---

[2] http://www.whitehouse.gov/blog/2013/08/06/incentives-support-adoption-cybersecurity-framework

[3] http://www.nia.nih.gov/about/events/2013/white-house-workshop-psychological-science-and-behavioral-economics-service-public

Framework of Analysis: Stakeholder Mapping

The first task to be addressed is to develop a framework to help identify those areas where the behavioral economics approach could be applied with the highest return on investment of effort. This requires a map of stakeholders. At a very high level the stakeholder environment can be divided into three broad areas:

- Major players, which include network operators, Fortune 500 enterprises, federal agencies, etc. These players are sophisticated in cybersecurity systems and practices (they understand the scale and importance of the problem), they spend significant dollars, and work through hierarchies based on relatively clear-cut lines of authority.

- Entities further down the value chain, including local service providers and partners (e.g. MSPs), tier two/tier three supply chain enterprises, state and local governments, etc. These stakeholders have a critical role to play, but are sometimes much less focused on cyber security (which is not seen as central to their mission) and often have relatively fewer resources to dedicate to supporting cybersecurity.

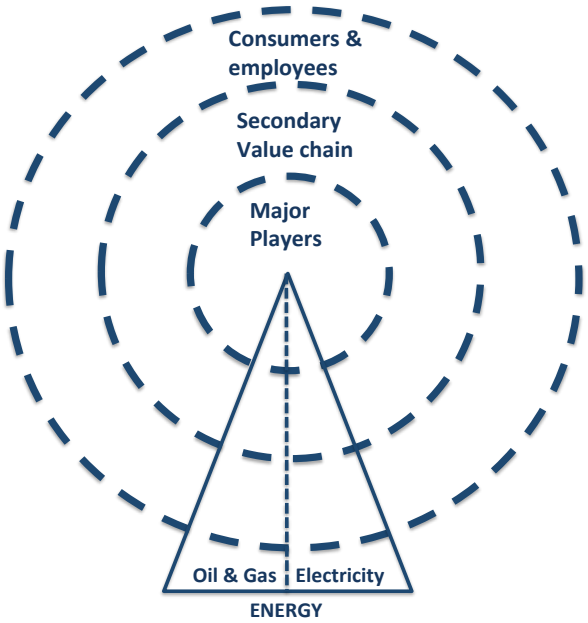
- Consumers and employees (often one and the same) are the foundation for any comprehensive approach to improving cybersecurity. There is dramatic variation in their level of knowledge of cybersecurity practices, and in their willingness to adhere to good practice either at home or in the workplace.

As a result of technological change and the disintegration of corporate and governance networks these three areas are thoroughly blended. This underscores the extent to which long-term improvements in cybersecurity will require a wide range of approaches across many domains involving all stakeholders. There are no silver bullets. However this map of stakeholders can be used to identify particular, higher-priority areas for intervention that are substantively important and which will make a measurable contribution to cybersecurity within a specific policy domain.

Stakeholder Mapping Example: Power Generation & Distribution

Major stakeholders often have the centralized authority necessary to require behavior by corporate entities, employees, and citizens. They also have the resources and technical capacity to support the implementation of such requirements. Behavioral economics can certainly be useful for such stakeholders, but rules, regulations and resources are obvious first steps. In contrast, smaller entities, further down a supply chain or with limited jurisdiction, have less authority and fewer resources. As it happens, such entities play a critical role in one of the nations' key infrastructures—the electric power grid. While the major players, such as Investor Owned Utilities (IOUs), have resources and technical capacity, their smaller supply-chain partners have come to the challenge of cybersecurity late in the game.

Small and Medium Enterprises (SMEs) who supply power systems equipment have been accustomed to designing and using control systems that were literally cut-off from the wider world by fences and locked doors. But the world has changed. With the development of distributed management of the grid and the increased use of outsourcing by IOUs, these power systems SMEs represent a soft spot in the web of cybersecurity measures necessary to protect the power systems infrastructure. Furthermore, behavioral economics could be especially useful in this area because of its value in solving information problems—establishing the importance of cybersecurity—among a population for which it is a low priority, and in making effective use of limited resources in the absence of controlling authority—using choice architecture and other devices to encourage actions through persuasion and at low cost.



In summary, the secondary value chain within electric power sub-sector is an area ripe for the application of behavioral economics because:

- This sector is a critical national infrastructure
- Players in the secondary value chain are vital to securing critical infrastructure, but relatively slow to adopt better cybersecurity practices
- These actors can benefit significantly from policy approaches that offer low-cost solutions to information problems and incentive structures in the absence of hierarchy.

This is not to overlook the importance of consumers and employees. Every entity from major stakeholders downwards all employ individuals who are both, and whose low levels of information and weak incentives for good practice represent another key weak-spot in the web of cybersecurity. As the discussion that follows will show, all of the foundational concepts in behavioral economics apply strongly to individuals, because behavioral economics is designed to amend and improve a standard (neo-classical) micro economic framework.

Government policy and incentives affect the decisions by actors at every level of this framework. In the energy sector, public utilities make investment decisions based in part on the extent to which those investments can be recovered through inclusion in the "rate base" for those entities. As pointed out by one speaker at the 2013 colloquium following the Workshop on the Economics of Information Security, investor-owned utilities tend to underinvest in cybersecurity as state regulators do not make such investments "allowable" under rate regulation. At the consumer level, acceptance of "smart meters" has varied based on how the benefits are explained to ratepayers, and the level of sentiment about smart metering could invade privacy by collecting additional data on household habits.

The energy sector illustrates the pervasive influence of government policy, and how it may have contradictory results. Regulation, legislation, and policy shape the stakeholder framework by establishing the environment in which each set of stakeholders makes choices, and how they relate to one another. This stakeholder framework highlights how government action on incentives must differ for each stakeholder group, and how the response to incentives may vary based on the nature and positioning of each group.

Insights from Behavioral Economics on Socio-Economic Incentives

Behavioral economics still has a reputation among non-experts as a relatively novel and untested approach. However, a number of core insights and findings of experimental economics have been replicated, extended and subject to rigorous discussion in the academic literature over several decades. (The approach may be said to have entered the mainstream when an early and distinguished researcher in the field, Daniel Kahneman, received the Nobel Prize in 2002). These have been the basis for policy proposals and actual interventions in other domains,

and so provide credible examples of the kinds of applications worth emulating in the case of cybersecurity.

The table below offers up descriptions of key elements in this approach that are generally accepted (this is not an exhaustive list). Each element is described in terms of the standard microeconomic approach and that of behavioral economics. The heart of the differences between these approaches is an assumption about how individuals choose. The neo-classical microeconomic approach assumes agents are well informed, or capable of gathering all necessary information, capable of calculating payoffs in an unbiased way, and capable of using the results as the basis for rational choice. The behavioral approach assumes that humans have enduring tendencies and biases that shape actions in ways inconsistent with so-called "rational man".

	Standard Microeconomics	Behavioral Economics
Loss	Humans treat the prospect of a loss the same as the prospect of a gain when each are of equal value	Humans are risk averse when faced with a loss and risk accepting when faced with a gain when each are of equal value
Fairness	Humans do not allow ideas about fairness to influence transactions, and always accept any offer that improves their payoff	Humans have evolved to care strongly about fairness, and punish others in transactions thought to be unfair even if it reduces their payoff
Context	An individual's preferences are the same no matter how they are framed or in what order choices are presented	An individual's preferences can be shaped by framing (the context) and ordering choices in different ways
Confirmation	All other things being equal, individuals weigh old and new information equally when making judgments	Individuals weigh new information more heavily when it is consistent with old information and existing beliefs
Peers	People choose based on the payoffs they will receive, independent of the choices of others	People are susceptible to peer pressure and make choices that match the choices of neighbors
Policy	Individuals will conform to policy when the negative consequences of noncompliance (e.g., penalties) outweigh the costs imposed by compliance.	Individuals make differing judgments about the extent to which they are willing to risk "getting caught" by an authority in order to avoid the sacrifices required by compliance.

Examples of behavioral change required to promote sound cybersecurity practices can be divided into three broad categories:

- **Information:** How to credibly share information among entities and individuals so that it is easier to choose wisely
- **Adoption:** How to change cybersecurity practices and policies currently in use
- **Maintenance:** How to ensure the continued use of sound cybersecurity practices and policies.

Other domains of public policy have instructive findings for each of these areas:

Context, confirmation and information: The way information is shared and framed is critical to the likelihood of it being noticed, believed and acted on. For example, which year you choose to claim social security between the years of 62 and 70 makes no difference from an actuarial perspective, but because 65 (at present) is described as the year when "full benefits" begin there is a spike in new claims for that year.[4] Other factors also influence how information is treated. Is it endorsed by a credible source and presented in a positive context as a worthy activity?[5] Is it consistent with our existing information (or "anchor")?[6] Providing information to consumers and employees about cybersecurity that frames it as a virtuous activity (people raise their level of performance in some cases when intrinsic rewards are emphasized), or as a familiar safety practice, like wearing a seatbelt, makes it more likely to foster the desired behavior.

Fairness and adoption: In the absence of affective monitoring by authority, having team members bear responsibility for sanctioning shirkers is an obvious approach. However, standard approaches assume that only small groups with excellent information can coordinate sanctions. In contrast, behavioral economics shows how some group members will pay to sanction shirkers.[7] We may imagine, therefore, that cybersecurity practices can be more easily adopted when the institutional environment allows altruists within the team to sanction others in order to raise the performance of the team as a whole.

Peers and adoption: In the absence of authority changing behavior is difficult and takes many years. The campaign against smoking is only now successful because many teenagers view it as uncool. This is a recent development, for decades the reverse was the case. However, the credible use of information can mobilize the impact of peer pressure. Hospitals forced to disclose infection rates are more willing to adopt new checklist systems that will improve their rankings.[8] Consumers of electricity are more inclined to practice conservation if they know how they rank compared to their neighbors, and if their good behavior is flagged on their monthly

[4] http://www.rand.org/pubs/working_papers/WR793.html
[5] http://www.nytimes.com/2005/05/15/books/chapters/0515-1st-levitt.html
[6] http://psych.cornell.edu/sites/default/files/Epley%26Gilo.06.pdf
[7] http://www.sciencedirect.com/science/article/pii/S0167268109000778
[8] http://www.newyorker.com/reporting/2007/12/10/071210fa_fact_gawande?currentPage=all

energy bill.[9] Peer pressure can influence entities and individuals whose performance is publicly shared with a wider community or whose aspiration to do the right thing is given a concrete metric.

Loss and maintenance: In a recent field experiment in education teachers were paid in advance but then required to repay some of their salary if their students performed badly. These teachers significantly out-performed the control group.[10] Imagine, instead, that managers were paid in advance for high-levels of compliance within their group on a cybersecurity practice, and had to give the money back if they fell short. Performance would almost certainly improve.

Policy. Examples of behavioral responses to government action abound in the policy literature. While motorcycle helmets are required in most U.S. jurisdictions, compliance still varies depending on the degree to which motorcycle riders perceive the relative balance between the penalties for non-compliance and the safety benefits of wearing helmets. This example also reveals a particular issue with incentives determined by the government, in that some motorcycle riders choose not to wear helmets precisely because usage is mandated by government. Advocacy groups, especially those focused on individual liberties and leaning towards libertarianism, have succeeded in repealing universal motorcycle helmet laws in some states. Their argument is that such laws are undesirable because they involve "paternalism" on the part of government—despite clear proof that these laws reduce fatalities and injuries from motorcycle accidents.[11] In this case, government incentives or mandates encounter resistance specifically by their governmental nature.

In summary, it is clear that well-established elements of the behavioral economics approach may provide the basis for interventions in support of cybersecurity practices and policies. They are especially helpful in environments where authority and resources are in short supply, such as the secondary value chain in the electric sector. Further, there is plenty of evidence that a sustained program of research mobilizing all the elements of behavioral economics (many not touched on here) behind potential policy interventions in cybersecurity would be a fruitful endeavor. The next step would be a comprehensive literature review to establish the baseline for existing knowledge and to identify promising applications across the cybersecurity policy space worthy of a systematic research program.

---

[9] http://www.forbes.com/sites/eco-nomics/2011/04/05/op-ed-behavioral-economics-and-your-monthly-energy-bill/2/

[10] http://www.nber.org/papers/w18237.pdf

[11] M.M. Jones and R. Bayer, "Paternalism & its discontents: Motorcycle helmet laws, libertarian values, and public health," *American Journal of Public Health*, 97(2), February 2007, pp. 208—217.

Timeline & Deliverables

This paper envisions a six-month research and planning process with two thrusts—one in formulating a research agenda in cybereconomics, and one in designing a field experiment ("proof-of-principle") in the application of behavioral economics to the design of incentives around cybersecurity practices. These activities and timeframes are shown in the table below.

Start Date	Activity	Deliverable	Due Date
9/1/2013	Revise and finalize concept paper (this document)	Final version of concept paper	12/17/2013
10/1/2013	Literature review on extant work in cybereconomics (drawing heavily from WEIS proceedings)	Literature review summarizing key research themes and highlighting gaps in the literature	1/3/2014
12/1/2013	Develop research agenda for future work on cybereconomic incentives	Proposed research agenda providing research themes and key areas of interest	1/31/2014
12/15/2013	Literature review on behavioral economics and incentives, focusing on its application in domains outside cybersecurity	Literature review and key findings from the application of behavioral economics and other social science to develop incentives modifying human and organizational actions	3/31/2014
2/15/2014	Design a potential field experiment, using the concept paper framework and findings from behavioral economics, to test a method for conducting field research in cybereconomic incentives	Proposed experiment, including research design, protocols, possible test population(s), research questions and hypotheses	4/30/2014

SRI International

Literature Review:
Current Research in Cybereconomics

January 3, 2014

Prepared for:
Dr. Joseph Kielman
Cyber Security Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Lucien Randazzese, Ph.D.
Center for Science, Technology & Economic Development
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
lucien.randazzese@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

1 Introduction

1.1 Motivation for this Review

Facing threats from cyber attacks that could disrupt our power, water, communication and other critical national systems, the President issued Executive Order (EO) 13636 on *Improving Critical Infrastructure Cybersecurity*,[1] and on the same day Presidential Policy Directive (PPD) 21 on *Critical Infrastructure Security and Resilience*.[2] These policy documents give the Department of Homeland Security (DHS) an overall coordinating role in pursing the cybersecurity objectives outlined in each document. Accordingly, an interagency task force led by DHS was established with eight working groups, one of which focuses on incentives. The working group was responsible for conducting a study of incentives for entities to adopt elements of the voluntary critical infrastructure cybersecurity framework developed by NIST as part of the requirements of the EO/PPD. The study of incentives, even with respect to their specific impact on cybersecurity, was broad in scope and diverse in focus of coverage.

In an effort to inform the development of a research agenda that will help DHS and the federal government understand how incentives can be used to improve overall national cybersecurity, SRI is conducting a two-part review of current research for the DHS Science and Technology Directorate (S&T) Cyber Security Division (CSD). The first part of the review, presented in this paper, surveys the extant research in cybereconomics. Much, though not all, of this research addresses incentives through the lens of traditional economics that assumes rational cost-benefit decision making on the part of companies, criminals, and the public.

In addition to this cybereconomic research there is a growing body of work that focuses explicitly on the behavioral aspects of human decision making, and the insights from this research has found its way into a variety of policy application areas, ranging from energy efficiency to public school teacher effectiveness. Cybersecurity is an obvious area in which a firmer understanding of human decision making in the face of uncertain payoffs and threats would be of great value, and research on this topic has indeed been growing in recent years. The second part of the literature review, to be provided in a subsequent paper, will focus on behavioral economics and other social science research primarily in domains outside of cybersecurity, with specific attention to real world applications of research findings to influencing incentives across a range of policy challenges.

The dichotomy between the research reviewed here and what will be presented in the second part of the review is not black-and-white. Over the last few years the community of scholars that has coalesced around the topic of cybereconomic incentives has begun to actively consider the behavioral aspects of cybersecurity, and the results of this behavioral research are included in this paper. Bear in mind that the focus on the second part of the literature review, to be provided in a subsequent paper, will focus primarily on the application of behavioral research in policy domains outside of cybersecurity.

1.2 Scope & Organization

This review of cybereconomics covers most of the primary work in the field, and draws extensively, though not exclusively, from the research presented over the years at the Workshop on the Economics of Information Security, or WEIS. WEIS is the leading forum for interdisciplinary scholarship on information and cyber security, combining expertise from the fields of economics, social science, business, law, policy, and computer science. A significant portion of the papers published on cybereconomics trace their roots to WEIS, and a very large share have been co-published there, or presented there as drafts prior to eventual publication.

Research in this review has been organized into three sections:

- Section 3: Models of Investment in Cybersecurity
- Section 4: Studies of Cybersecurity Related Behavior
- Section 5: Cyberinsurance and Cyber Liability Research

Section 4, the largest of the three review sections, divides research into modeling (theoretical) work (Section 4.1), empirical work (Section 4.2), and experimental work (Section 4.3).

Deemphasized in this review are studies of such things as advertising spam, click fraud by consumers, consumer ad avoidance and other activities that, while related to and sometimes associated with cybersecurity risk, are largely issues of internet marketing, be it formal corporate marketing or grey market marketing. For an example of this type of research, see Vratonjic et al. (2012) who develop a game theoretic analysis that pits digital content owners against consumers enabled with online ad avoidance software to determine optimal strategies for content monetization.[3]

A share of the research featured in forums dedicated ostensibly to the study of cybereconomics actually tackle questions that are technological in nature, and are thus not reviewed in detail here. For example, Ransbotham (2010) conducted an interesting empirical study on the diffusion and exploitation of vulnerabilities in open source versus closed source software that was presented at WEIS 2010.[4] Ransbotham's analysis indicated that open source software vulnerabilities are at greater risk of exploitation, and diffuse more rapidly. While this type of research is important to the cybersecurity community generally, it is not directly related to cybereconomics or incentives, and is not included in this review except in those cases where there is a clear economic or incentive component to the research.

Three current cybereconomic research projects supported by DHS S&T CSD are not explicitly reviewed in this document. The projects are midstream, so it is premature to review their results with any conclusiveness. However, the projects are similar in focus, methodology, or both, to research presented here. For context, these three projects primarily would be categorized in this review as follows:

- Carnegie Mellon University, *Understanding and Disrupting The Economics of Cybercrime* → Section 4.2.3, Attacker Behavior
- University of Maryland, *Cyber Economics* → Section 3.1, The Public Good Challenge of Cybersecurity, and Section 4.2.2, Organizational Behavior
- University of Michigan, *Towards a Global Network Reputation System: A Mechanism Design Approach* → Section 4.3, Experimental Evaluations of Behavior

2 Summary of Research Findings

The initial work in cybereconomics, which began to emerge in the early 2000's, was in the field of microeconomics, and featured theoretical mathematical models that highlighted the public good nature of investments in cybersecurity. Given the externalities inherent in protection systems, this work concludes that firms left to rely on their own private return on investment (ROI) for investment decisions will underinvest in cybersecurity. While this research provides occasional insight into the economic dynamics of security investment, it is often theoretical in nature, producing results that are either common sense in nature or vague in their implications for policy.

This microeconomic work also largely ignores differences between major players (such as large corporates and federal government agencies) who tend to invest considerable resources in cybersecurity and secondary players in the value chain who invest much less and are thus more vulnerable to attack. More generally, there is little consideration of just how far from optimal investment is overall, as a whole or on the part of individual entities. The policy implications of a nation that is grossly underinvesting in security are obviously different than those in the case where investment is close to optimal.

In reaction to the limitations of the theoretical research, research focusing on actual behavior – individual and organizational – has become a central focus in studies of cybereconomics in the last several years. This research includes some theoretical work, but also makes novel use of specific datasets (e.g. consumer website usage) to examine behavior empirically, as well as formal experiments on how people behave in various cybersecurity settings. Some of the most interesting cybereconomic work is taking place in this domain. For example, two sets of researchers have shown that people are less likely to behave offensively online when their actual identities are shown along with their comments. Similarly, a number of researchers have shown that people are more likely to divulge sensitive information when they are told other people have, or when they feel they are in control of how information is revealed, *even when that control has no impact on who ultimately can gain access to the information being divulged*. The implications of such findings for cybersecurity would seem clear.

Real world data has also been used to investigate the motives and activities of cybercriminals. Data from captured infrastructure used in the conduct of actual cybercrime shows that cybercriminals seek to keep scams going as long as possible, even to the point of issuing

refunds to complaining (i.e. scammed) customers in effort to avoid detection. This type of research represents a novel and potentially useful approach in analyzing cybercriminal motivations and behavior.

While behavioral research has made some headway into our understanding of the decisions people make with respect to sensitive information, most of it focuses on personal privacy issues, and less on how people behave when they serve as the stewards of someone else's sensitive information, for example as employees of companies or government organizations with knowledge of proprietary or classified data, or in control of critical cyber infrastructure. Employees, or course, represent a significant source of cyber risk, and so looking specifically at their behavior, and why they do or do not comply with security policies, would make a fruitful path of research. There is some preliminary research, reported on in this review, of the drivers of good security practice at the organizational level, but less so at the individual level. One interesting result on organizational behavior, one that shows organizations to be as subject to psychological biases as are people, is that organizations will try to reduce the amount of spam they are responsible for when their spam levels are publically reported, but also that they put less effort into this reduction when other groups are reported as considerably worse offenders.

The lack of data is a critical methodological handicap to cybereconomic research, one often lamented by researchers. Among the areas for which there is a need for more data are:

- Cybersecurity policies and activities
- The costs associated with cybersecurity
- Security incidents and their outcomes, both technical and non-technical

The disincentives for sharing data are well known, but the common benefits to more widely available information are also clear. Notwithstanding this need for data, only one paper reviewed in this document explored possible incentives for organizations to share data.

If the reader is interested, Anderson and Moore published a 2006 review on cybereconomic research in *Science*, "The Economics of Information Security,"[5] that focused specifically on information security and the economic challenges to adequate security investment.

3 Models of Investment in Cybersecurity

Some of the first research in the field of cybereconomics, and a large share of the extant literature, examines investment by organizations in cybersecurity from a classical microeconomic perspective. This research frames the objectives of various participants in the "cyber marketplace," including network owners as well as attackers who wish to gain from breaching security measures, in terms of models based on formal mathematical expressions. To ensure the often complex sets of mathematical equations in these models are solvable, researchers make simplifying, stylized assumptions about many aspects of cybersecurity

investment. These models may, for example, assume complete knowledge about potential attacks or defense, examine simplified duopoly industry structures, consider scenarios of a single attacker and single target, or consider only a single investment period.

While individual models may relax specific simplifying assumptions in order to extend the literature, this work remains fundamentally theoretical. It nevertheless provides high-level insight into the challenges in ensuring that those best in a position to protect critical cyber infrastructure invest sufficiently in security. The first of the following two subsections looks at the general economic challenges associated with investing in cybersecurity given its public good nature. Following that, in Section 3.2, research on security investment in certain specific situations is reviewed.

3.1 The Public Good Challenge of Cybersecurity

Explicit in much of this research is the externality-driven lack of incentives for private investment in cybersecurity. Individuals and especially organizations rarely bear the full consequences of successful attacks on their information infrastructure. Thus the private penalty for underinvesting in security is smaller than the total, public cost, and so investments in security are sub-optimal from a social welfare perspective. Anderson and Moore (2006) discuss how misaligned incentives between those responsible for security and those who benefit from protection are rife in IT systems.[5]

The first significant attempt to examine economic incentives to invest in information security is an influential paper by Gordon and Loeb (2002).[6] The authors create a model in which information owners can invest to reduce the likelihood that an attack is successful but not to reduce the likelihood of the threat itself. They further assume that the security investment smoothly translates into security, without any lumpiness in required expenditure, and that potential attackers do not change their behavior in response to security investments by targets (i.e. the model is one-period). Based on these restrictive assumptions, Gordon and Loeb show that the optimal investment in information security is an increasing function of the vulnerability of information (vulnerability defined as the probability that an attack would be successful). Their analysis also shows that when the cost of protecting information gets sufficiently high, it ceases to be economically optimal to invest in its protection. Other researchers have used this paper as a basis for further research, cf. Tanaka et al. (2005), who show empirically (in the case of local governments in Japan) that organizations do indeed invest in security depending on the vulnerability of information;[7] Tatsumi and Goto (2009), who extend the model using options theory to examine how vulnerabilities affect the timing and amount of investment;[8] and Baryshnikov (2012), who looks theoretically at the generalizability and limitations of the Gordon and Loeb approach.[9]

Anderson (2006) described the information asymmetry problems associated with investment in cyber security.[10] The security market is subject to inaccurate disclosure, where organizations

are incented to underreport damage from cyberattacks for reputational reasons, while vendors of security technology are incented to exaggerate the risk of attack. The market for protection is also subject to "market for lemons" conditions in which much of the security technology available is of low quality because there is no effective way to demonstrate quality to potential buyers. This creates low confidence in security products generally. It thus attracts a disproportionate share of low quality products because providers of high quality products will be unable to earn a return on that quality.

Closely associated with the externality problem is the free rider problem, in which owners of information assets are positively affected by investments in cybersecurity by other asset owners, and thus are incented to invest less themselves. In another influential paper, Varian (2004) used game theory to illustrate that because harm can spread well beyond the network of an attacked entity, security investment is a public good, and some individuals may tend to shirk on their responsibility to secure their own assets.[11] Under such circumstances there will be underinvestment in cybersecurity. Several researchers have extended Varian's original model, to examine specific cybersecurity attack and defense situations. For example, Grossklags et al. (2008) look at various scenarios (e.g. attackers who only attack poorly defended targets), and evaluate theoretically how the socially optimal level of security investment would depart from the modeled level of investment in these situations.[12]

3.2 Diving Deeper Into the Challenges of Investing in Cybersecurity

In addition to research that highlights the theoretical reasons why there is likely to be underinvestment in cybersecurity overall, a number of researchers have started to examine more tangible and specific issues associated with investing in cybersecurity, and move toward somewhat more prescriptive research. For example, Hofmeyr et al. (2011) conducted simulations of autonomous systems within the Internet, where an autonomous system corresponds roughly to an ISP, in order to evaluate the most cost-effective way to combat malware across the entire Internet.[13] The authors' modelling shows that the best overall internet security, given a limited security budget, is achieved by intervention on a small group of the largest autonomous systems. Their analysis indicates that attention to the largest 0.2% of autonomous systems is more effective than improving security for a randomly chosen subset of 30% of all systems.

Brecht and Nowey (2012) provide a thorough review of the many challenges faced in quantifying enterprise cost for information security in a comprehensive fashion, challenges that make it harder for firms to evaluate and optimize their investment decisions.[14] After identifying these challenges, the authors go on to discuss various approaches for categorizing and determine calculating security costs in an enterprise.

While the shift toward research with more prescriptive potential is a positive sign, much of this more applied, micro-focused work is obvious in its conclusions. For example, Cezar et al. (2010)

look at how contracts can be constructed to ensure the optimal mix of outsourcing for security device management and for security monitoring;[15] the authors' proposal ultimately depends on cost complementarity between these two activities. Similarly, Ioannidis et al. (2011) use utility theory to examine organizational decisions to defer known-cost security investments,[16] such as security patch implementation, and illustrate how these decisions vary depending on the availability of resources and the opportunity costs of their implementation.

4 Studies of Cybersecurity Related Behavior

Moving beyond identification of the theoretical and practical challenges associated with achieving optimal levels of cybersecurity, there is a growing body of work that looks at how individual entities actually behave within the cyber arena. This research includes examination of the behavior of organizations, network owners, cybercriminals, and consumers. From an analytical perspective, this research is divided below into three broad categories of investigation: the economic modeling of behavior; empirical evaluations of behavior; and experimental evaluations of behavior. The experimental research does produce empirical results, of course, but is nevertheless addressed separately as it represents a methodologically distinct and promising avenue of research.

4.1 Economic Modeling of Behavior

In work that is methodologically similar to that described in Section 3.1, many research papers approach the assessment of behavior using theoretical work or mathematical modeling. This work focuses at a more granular level than that described above in Section 3.1, and addresses behaviors with respect to specific situations, technologies, types of organizations, etc. For example, August et al. (2013) assess how patching behavior changes as software applications migrate from on-premises to the cloud via Software as a Service (SaaS) applications.[17] Looking at corporations, Sapi and Suleymanova (2013) examine firm incentives to acquire customer data based on data quality and heterogeneity in customer characteristics.[18] And Khouzani et al. (2013) model the incentives for Internet Service Providers (ISPs) to adopt intrusion detection and prevention systems, and show how policymaking around adoption is complicated by free rider effects and variation in technical performance levels across ISPs.[19]

While this research is ostensibly aimed at understanding drivers of the behavior of various actors in the cyber ecosystem, and ultimately at informing public policy and management, it is often frustratingly theoretical. The "insights" from this work, while deriving neatly from elaborate mathematical formalism, don't always square with common sense expectation, or are often too vague or dependent on assumptions to be of substantial use in policy. For example, Zhou and Johnson (2009) model the impact of information security ratings that disclose information risk among network partners, and find that ratings can hurt both high-security and low-security providers, or benefit both, depending on the market conditions.[20] In another example, Baldwin et al. (2011) show that security professionals are inclined to make infrastructure decisions solely

on technical bases, but when enabled to make decisions using both technical and economic criteria, they make a different set of decisions.[21] While this result is interesting, and has some intuitive appeal, it is not clear which set of decisions is actually better, especially in the presence of known incentives to underinvest in cybersecurity (see Section 3.1 above).

Potentially more helpful is research like that of Ioannidis et al. (2013) who model the role an information steward has on information system defensive expenditure.[22] The authors show that in the absence of a steward – who is formally responsible for protecting an organization's information – individual actors will underinvest in expenditures because they will underestimate the expected losses from attacks. Conversely, in the presence of a steward, firms will spend more, significantly enhancing system sustainability.

While perhaps somewhat insightful, the Ioannidis et al. (2013) paper highlights another shortcoming of much of this work, namely, when its results are prescriptively valuable they also seem fairly obvious. Consider the following several papers.

Narasimhan et al. (2010) use game theory to examine the conditions in which network operators are likely to cooperate in making network security investments,[23] as it is widely acknowledged that internet security issues can be handled better through cooperation (again, because of the public good nature cybersecurity). Their findings indicate that incentives for joint security investments depend on player views on attack likelihood and losses. In Böhme and Moore's (2009) model of security investments the authors show that the best strategy depends on the defender's knowledge of prospective attacks and the sunk costs incurred when upgrading defenses reactively.[24] And Kolfal et al. (2010) examine firm willingness to spend on security in the face of competition for customers who can be adversely affected by security failures, and show that security investment varies depending on customer response to such adverse events.[25]

In similar work, Wellman et al. (2013) use simulation in a game theoretic approach to assess incentives for network protocol compliance,[26] and find that the benefits of complying with protocols are particularly strong for nodes (within a cyber network) subject to attack. This research highlights a situation in which the externality problem of cybereconomics effectively vanishes – when an entity views itself as vulnerable to a sufficiently costly attack. But again, this result follows from simple logic.

Moore et al. (2010) scale up the unit of analysis and use game theory to assess the cybereconomic behavior of nation states.[27] The authors consider nations that are political rivals, and describe scenarios in which they may be incented to 'stockpile' (keep secret their knowledge of) discovered security flaws, at the expense of improving security of civilian computer networks, in order to ensure military network superiority over their international rivals.

Not all the theoretical research on cybereconomic behavior is unclear in its implications or overly obvious in its conclusions. Some newer research, while still theoretical, is increasingly practical in its focus. For example, Kelley and Camp (2012) use epidemic modelling to assess optimal patching behavior with respect to reducing system-wide vulnerabilities that arise due to unprotected computers.[28] They show that small increases in patch rates and recovery speed can be very effective in reducing system wide vulnerabilities. If this result is borne out by further research, it implies a high ROI for patching.

Similarly, Thomas et al. (2013) propose a branching activity model to estimate the impact of information security breaches.[29] The authors measure the impact of a breach in terms of the combined efforts that all affected stakeholders would be willing to spend to recover from the breach. They describe an approach for estimating the parameters used in the branching model using primarily public source "Indicators of Impact," such as news reports and regulatory filings that report on cybersecurity events. This approach, in theory, could have use in enterprise risk management, real-time crisis management, and resilience planning, among other areas.

These recent papers highlight a trend in cybereconomic research; some of the most proscriptively valuable papers address best practices in network management, and can only indirectly be considered focused on cybereconomics per se.

4.2 Empirical Evaluations of Behavior

Given the limitations of theoretical modeling in informing policy on cybersecurity, a growing body of research is studying cybereconomic behavior from an empirical perspective, using novel sets of data on actual behavior – individual and institutional – to shine light on how people and organizations *actually behave* with respect to the many security issues that surround cyber assets. Some of the data researchers have been using in these studies actually come from formal experiments of individual behavior, conducted in controlled and usually synthetic situations. This experimental research is reviewed separately in the next section (Section 4.3). Interestingly, some topics have received both empirical and experimental considerations – for example consumer willingness to disclose sensitive information online.

The review that follows organizes cybereconomic behavioral research into that focused on individual behavior, organizational behavior, and attacker behavior.

4.2.1 Individual Behavior

While the arms race between critical infrastructure owners and those who would attack their information asset continues, effective cybersecurity remains as much a human issue as it does a technological one. Reflecting this emphasis on human factors, much of the empirical cybereconomic research looks at how people behave when dealing with sensitive electronic information.

Takemura and Komatsu (2012) used survey data to assess employee violation of information security rules.[30] The authors looked at a variety of individual and workplace factors, and find that information security violations are correlated with more cavalier attitudes towards risk, low information security awareness, and permanent employment structures, but are not correlated with workplace satisfaction. Employees obviously represent a significant threat source for data breaches and other cybersecurity failures, and so understanding what makes them comply (or fail to comply) with policy has important implications for improving security.

Notwithstanding the importance of employees in the cybersecurity ecosystem, as a group there has been less research focused on employees than on consumers and how consumers behave with respect to their own data. Cho and Acquisti (2013) examined over 75,000 comments on online news articles and show that commenters are less likely to make offensive comments when their true identities (rather than pseudonyms) are presented with comments.[31] Similar research by Cho (2011) examined the impact of the Real Name Verification Law implemented in Korea that requires use of real names in most online discussion.[32] Cho's analysis showed that the law reduced offensive commenting, but also that it did not affect online discussion participation in the long term (though there was a measureable short term effect). This research has clear implications for those trying to improve cybersecurity, as it strongly suggests that people will behave more socially acceptably and responsibly when they cannot hide. It's easy to imagine how knowledge of this behavior, if confirmed in further study, might be applied to human factor considerations in security policies design.

In similar work, Cavusoglu et al. (2013) examined panel data from a popular (unnamed) social media site to explore how privacy settings affect user information disclosure behavior.[33] The authors find that more granular control leads to more open disclosure. A related study by Tucker (2011) shows that consumers are more likely to respond to (less likely to object to) internet advertising that has used their personal information in targeting and content when users have more granular control over what personal information is shared on social media sites, *even when this control has no effect on who ultimately has access to that information*.[34] These studies strongly suggest that individuals are more likely to divulge information when they believe they have more control over that information, regardless of whether that control actually impacts use of data by others.

Preibusch and Bonneau (2011) look empirically at consumer willingness to pay for privacy in online consumer services, including both paid and free services.[35] The authors looked at 140 websites and find that differentiation in privacy polices across websites providing competing services is more prevalent for priced (versus free) products, but surprisingly, sellers that collect fewer data charge lower prices.

Wash (2010) conducted a qualitative study of home computer owners to understand their mental models of attackers and home security technology, what he refers to as folk models, to understand how these models affect their home security decisions.[36] Wash finds that

consumers preconceived ideas of cyber threats determined how securely, or insecurely, they behave. If nothing else, this research illustrates the importance of understanding the preconceived notions of non-experts when considering how they will behave with respect to sensitive information.

Neuhaus and Plattner (2012) look at vulnerability fix behavior for reported vulnerabilities for open source software (Mozilla, Apache, and Tomcat) in an effort to determine whether, for this software set at least, the point of diminishing marginal returns to fixing has been reached.[37] Their study does not investigate behavior per se, but uses that behavior as a way to measure the economics of vulnerability fixing. The authors show empirically that fix rates are not declining, suggesting that diminishing margin utility to fixing has not yet been reached for this software set. Neuhaus and Plattner had expected to see diminishing rates of fixing, as theory suggests that fixes are increasingly hard to achieve over time as the easiest vulnerabilities are addressed first. This result, if borne out by work in other areas, has implications for resource allocation, and assumptions made about the ROI on cybersecurity investments made at increasing levels.

Not all of the empirical research is as insightful as these studies. Some of it shares characteristics of the theoretical work in that its conclusions seem obvious. For example, Wood and Rowe (2011) show via analysis of survey data that consumers are willing to pay for security if that security leads to fewer negative events (e.g. computer crashes, identity theft).[38]

4.2.2 Organizational Behavior

Turning to consideration of how organizations behave, the healthcare industry has received a lot of the early research attention since it must deal not only the usual challenges of keeping information assets safe, but also must comply with high-profile and highly restrictive policies meant to protect consumer health information.

Kwon and Johnson (2011) examined the effectiveness of cybersecurity investments in the healthcare industry,[39] and show that proactive security investments are more effective than reactive ones in preventing future data breaches, and suggest a role for policymakers in influencing security investment decisions via regulation. Gaynor et al. (2012) considered the effects of competition on hospital data protection,[40] and find that hospitals in competitive markets actually perform more poorly in protecting patient data. The authors posit that in more competitive environments hospitals shift resources towards activities more visible to consumers; data protection not being one of these activities. This result, if confirmed, is somewhat counter-intuitive, and highlights the need to better understand how firms make cybersecurity investments in the context of tradeoffs made about spend in other areas. Closely related work by Appari et al. (2009) examines HIPPA compliance at acute care hospitals.[41] They find that privacy compliance is positively influenced by the comprehensiveness of state-level privacy regulations, current security compliance level at a hospital, employment of a dedicated

compliance officer with a compliance background, and the average level of HIPAA compliance of all institutions within the state in which a hospital resides.

Kwon and Johnson (2012) also looked at how organizational characteristics impact security performance and data-security related regulatory compliance in the healthcare industry.[42] They looked specifically at hospitals and measured compliance in terms of health data-related regulations such as HIPPA, and find that organizations with greater resources, greater security expertise, and top management who advocate for information security have fewer breaches and are more compliant. Though these results largely confirm a priori expectations, they do underscore the fact that security can be improved at the organizational level when organizations commit to its importance.

Looking at organizations more broadly, Asghari et al. (2013) examine the incentives of the actors in the security value chain, particularly the certification authorities.[43] Looking at certificate pricing, the authors find that certificate pricing is not driven by the sale of the certificates themselves, but by the services and reputation signals bundled along with the certificates. As a result there are large price differences across suppliers of what is essentially a commodity – certificates are certificates. This result suggests that security buyer behavior is driven as much by perceptions as it is by technical outcomes.

Wang and Kim (2009) examined behavior at the level of nation states; specifically the impact of joining the international Convention on Cybercrimes.[44] The authors' analysis indicates that participation in the convention subsequently reduced the number of attacks originating from signatories by 16% to 25%, a significant outcome. Recall that Kwon and Johnson (2012) showed that organizational commitment to cybersecurity has a positive impact on security outcomes. The Wang and Kim (2009) study shows a similar effect for nations; there are measurable positive security impacts when countries make a commitment to cybersecurity.

Hunker and Probst (2009) examine an interesting empirical observation at the interface between organizational behavior and individual behavior.[45] They illustrate that firms almost always take basic steps to prevent routine insider (employee) attacks, but do not typically attempt to address the potential of more serious attacks from insiders. The researchers' conclusion, in essence, is that the range of possible insider threats is too broad to mitigate cost-effectively ex ante. As a result firms simply ignore them. While this result is discouraging, it highlights again the importance of the human dimension in cybersecurity.

4.2.3 Attacker Behavior

Some of the most interesting empirical research in cybereconomics has been focused on cybercriminal and attacker behavior. Some of this research is made possible because researchers gain access to specific databases that contain data relevant to some particular aspect of cybersecurity. For example, Stone-Gross et al. (2011) examine the economics of fake antivirus software providers by looking at data stored on servers that had been (unwittingly)

used to facilitate cybercrime.[46] Analysis of these data shows that cybercriminals seek to keep scams going as long as possible, even to the point of issuing refunds to complaining (i.e. scammed) customers in an effort to avoid detection. Baldwin et al. (2012) provide empirical evidence that cyber threats to critical services (such as email, databases, name and directory servers, website operations, and shared storage) are correlated with one another, a pattern the authors refers to as contagious inter-relationships.[47] The authors highlight the implication that failure to appreciate these interrelationships might lead IT managers to underinvest in security, a theme consistent with the theoretical literature.

Research by Herley (2010) and Florencio and Herley (2011) seeks to explain why, when so many information technology users are known to use very poor security protocols, there are not more attacks on their information. This research is not, strictly speaking, empirical. Rather, it uses the same economic theory described in Section 3.1 to explain an important empirical pattern. Florencio and Herley (2011) show that the expected value of attacks is negative in a wide range of real-world circumstances, and thus explain the low observed attack frequency.[48] Herley (2010) explains that given the skewness in target value, cyber attacks are often focused on targets with high value in order to be profitable in expectation, and thus most users are never attacked.[49] Using actual market pricing available on internet relay chat networks for stolen identities, phishing kits, botnets, and cybercrime related services, Herley and Florencio (2009) show that cheating among cybercriminals dramatically reduces the profit from identity theft and related crime, but also contributes extensively to the externalities of cybercrime.[50]

Segura and Lahuerta (2009) use analytical and simulation modelling, in combination with data collected by the authors on actual prices charged by cybercriminals to launch DDoD attacks, to examine the incentives for DDoS attacks aimed at extorting their victims.[51] The authors postulate that with enough such pricing data (theirs was limited), one could estimate the probability of attack on a given service. More work is required to support this claim, but the research, along with that of Florencio and Herley, represents a fruitful methodological push for cybereconomics research, one that looks at real markets for cybercrime as a way to inform security and policy.

4.3 Experimental Evaluations of Behavior

Paralleling the empirical work described above, several cybersecurity researchers have started to investigate behavior using formal experiments. The subjects for the formal experiments are typically college students at the universities where the professors conducting the research reside.

Using just such a set of student test subjects, Acquisti et al. (2009) show that people are more willing to divulge sensitive personal information when told that other respondents have made sensitive disclosures.[52] The authors also show that people anchor their views of how intrusive requests for information are based on early or initial experience with those requests, and are

therefore more likely to disclose information when later requests for information are relatively less intrusive than earlier ones, independent of the absolute level of sensitivity of the information being requested.

Preibusch (2013) used an experimental setting to examine web search users willingness to use and pay for privacy options in search.[53] He finds that users are more likely to use privacy options when searches are more sensitive, but most users are not willing to pay even a small fee for their use. In related research, Preibusch et al. (2012) looked at consumer web form disclosure behavior, and found that consumers tend to over-provide sensitive information even when it is not required.[54] In other words, the simple act of soliciting sensitive data, even while explaining it was not required, was sufficient inducement for many consumers to provide such data. Similar research by Malheiros et al. (2012) showed that people are willing to answer questions on a credit card application that are highly personal and not connected to credit underwriting, such as, "*Did any of your loved ones die while you were growing up?*"[55] Taken together these experiments strongly indicate that people tend to provide sensitive information when asked for it, or when they perceive its disclosure to be part of some standard or accepted practice.

Brandimarte et al. (2010) explore another tendency with respect to disclosure, and show that people are more willing to disclose private information online when they have more control over the publication of that information, even if it such control has no effect on whether strangers can access the same information.[56] This result is very similar to the results of Cavusoglu et al. (2013) and Tucker (2011), who show empirically (see Section 4.2.1) that when people feel they are in more control of the disclosure of their data they tend to disclose more, even when that control is not connected to who ultimately gains access to the data.

Egelman et al. (2012) found that under certain conditions, consumers are willing to pay a premium for smartphone apps that make less intrusive information requests.[57] In another experiment by the same researchers, Egelman et al. (2010) found that individuals are more likely to tolerate online activity delays when told that the delays were for online security purposes and when the nature of the security threat was explained.[58] These experiments suggest that people are willing to pay for privacy, and are willing to tolerate more "hassle" in the name of cybersecurity.

Finally, using a quasi-experimental approach, Tang et al. (2013) evaluated the impact of public reporting of organizations' spam generation.[59] The authors report that organizations reduced outgoing spam by approximately 16% when subject to reporting. They also find that organizations do less to reduce their own spam when the worst-reported offender produces a greater absolute level of spam, indicating a social comparison effect.

5 Cyberinsurance and Cyber Liability Research

For a given organization, a high damage cybersecurity event is a low-odds, high-cost event, and thus characteristic of the type of events organizations protect themselves against with insurance. But despite commercial, academic, and policymaker interest in the prospect of cyberinsurance as a tool for addressing the risk of cyberattacks, the market for actual policies remains very underdeveloped. The failure of the cyberinsurance market to materialize has naturally become the focus of cybereconomic research. Using standard microeconomic theory, researchers have attributed the lack of a cybersecurity market to both the interdependence of security across insurance holders, cf. Bolot and Lelarge (2008),[60] and correlation of risk across insurance holders, cf. Böhme and Kataria (2006).[61] Information asymmetries have also been used to explain the lack of a cybersecurity market. For example, Shetty et al. (2009) model simplified, non-real-world scenarios in which insurers have either zero information or perfect information on user security, and show how insurance markets either fail to materialize or fail to improve security, respectively.[62]

Böhme and Schwartz (2010) provide an economic overview of various market models of cyberinsurance, discuss the variety of economic challenges each of these models would face, and ultimately suggest that cyberinsurance markets, despite the informal arguments made in favor of their use, may not be viable given the identified economic challenges.[63]

Examining a different aspect of cyberinsurance, Innerhofer-Oberperfler and Breu (2009) consider rating indicators for cyberinsurance that could be used to calculate premiums.[64] Given the lack of data on actual losses, the authors proposed a set of preliminary indicators, and assess which of these potential rating variables might best reflect actual risk exposure.

Fryer et al. (2013) examine legal liability theory for its application to cybersecurity, and the impact of liability on different actors in the internet security area.[65] The authors conclude that although there are instances where liability could have positive economic effects or provide protection to consumers (through effective internalization of externalities), in general the costs associated with administrating such a liability system and the practicalities of identifying victims and losses make cyber liability impractical. In this regard, the concept is similar to cyberinsurance in that it has intuitive appeal but may face practical limitations in application.

August and Tunca (2011) examine optimal liability regimes with respect to a specific case, namely software vulnerabilities.[66] The authors compare the effectiveness of vendor liability for customer damages, vendor liability for customer patching costs, and government imposed security standards. The research approach for this work is again theoretical economic modelling, and so it is not surprising that the results, while sensible, do not shed extensive light on how liability might be used to improve cybersecurity. Specifically, the authors find that the optimal liability regime depends on tradeoffs in attack likelihood and patching cost, a somewhat common sense conclusion.

Research Referenced in this Review

[1] The White House, *Improving Critical Infrastructure Cybersecurity*, Executive Order 13636, February 12, 2013.

[2] The White House, *Critical Infrastructure Security and Resilience*, Presidential Policy Directive 21, February 12, 2013.

[3] Vratonjic, Nevena, et al. "Ad-blocking Games: Monetizing Online Content Under the Threat of Ad Avoidance." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[4] Ransbotham, Sam. "An Empirical Analysis of Exploitation Attempts based on Vulnerabilities in Open Source Software." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[5] Anderson, Ross and Moore, Tyler. "The Economics of Information Security." *Science* 314 (2006), 610–13.

[6] Gordon, Lawrence M and Loeb, Martin P. "The Economics of Information Security Investment." *ACM Transactions on Information and System Security* 5.4 (2002): 438-457.

[7] Tanaka, Hideyuki et al. " Vulnerability and information security investment: An empirical analysis of e-local government in Japan." *Journal of Accounting and Public Policy* 24 (2005): 37–59.

[8] Tatsumi, Ken-ichi and Goto, Makoto. "Optimal Timing of Information Security Investment: A Real Options Approach." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[9] Baryshnikov, Yuliy. "It Security Investment and Gordon-Loeb's 1/e Rule." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[10] Anderson, Ross. "Why Information Security is Hard—An Economic Perspective." *Proceedings of the 17th Annual Computer Security Applications Conference* (2006), 610–13.

[11] Varian, Hal. "System Reliability and Free Riding." *Economics of Information Security, vol. 12.* Camp, L. J., Lewis, S. (Eds.). (2006): 1-15, Kluwer Academic Publishers.

[12] Grossklags, Jens et al. "Secure or Insure? A Game-Theoretic Analysis of Information Security Games." *Proceedings of the 17th International World Wide Web Conference* (2008), 209–218.

[13] Hofmeyr, Steven, et al. "Modeling Internet-Scale Policies for Cleaning up Malware." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[14] Brecht, Matthias and Nowey, Thomas. "A Closer Look at Information Security Costs." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[15] Cezar, Asunur, et al. "Outsourcing Information Security: Contracting Issues and Security Implications." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[16] Ioannidis, Christos, et al. "Fixed Costs, Investment Rigidities, and Risk Aversion in Information Security: A Utility-theoretic Approach." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[17] August, Terrence, et al. "Cloud Implications on Software Network Structure and Security Risks." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[18] Sapi, Geza and Suleymanova, Irina. "Consumer Flexibility, Data Quality and Targeted Pricing." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[19] Khouzani, MHR, et al. “Incentive Analysis of Bidirectional Threat Filtering in the Internet.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[20] Zhou, Zach and Johnson, Eric M. “The Impact of Information Security Ratings on Vendor Competition.” Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[21] Baldwin, Adrian et al. “Economic methods and decision making by security professionals.” Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[22] Ioannidis, Christos, et al. “Sustainability In Information Stewardship.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[23] Narasimhan, Harikrishna, et al. “Towards a Cooperative Defense Model Against Network Security Attacks.” Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[24] Böhme, Rainer and Moore, Tyler. “The Iterated Weakest Link, A Model of Adaptive Security Investment.” Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[25] Kolfal, Bora, et al. “Market Impact on IT Security Spending.” Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[26] Wellman, Michael, et al. “Analyzing Incentives for Protocol Compliance in Complex Domains: A Case Study of Introduction-Based Routing.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[27] Moore, Tyler et al. “Would a ‘Cyber Warrior’ Protect Us? Exploring Trade-offs Between Attack and Defense of Information Systems.” *Proceedings of the 13th New Security Paradigms Workshop* (2010), 85-94.

[28] Kelley, Timothy and Camp, L. Jean. “Online Promiscuity: Prophylactic Patching and the Spread of Computer Transmitted Infections” Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[29] Thomas, Russell Cameron, et al. “How Bad Is It? A Branching Activity Model to Estimate the Impact of Information Security Breaches.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[30] Takemura, Toshihiko and Komatsu, Ayako. “Who Sometimes Violates the Rule of the Organizations?: Empirical Study on Information Security Behaviors and Awareness.” Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[31] Cho, Daegon and Acquisti, Alessandro. “The More Social Cues, The Less Trolling? An Empirical Study of Online Commenting Behavior.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[32] Cho, Daegon. “Real Name Verification Law on the Internet: A Poison or Cure for Privacy?” Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[33] Cavusoglu, Huseyin, et al. “Do Privacy Controls Influence Content Generation and Sharing Patterns of Online Social Network Users? A Natural Experiment.” Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[34] Tucker, Catherine. "Social Networks, Personalized Advertising, and Perceptions of Privacy Control." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[35] Preibusch, Sören and Bonneau, Joseph. "The privacy landscape: product differentiation on data collection." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[36] Wash, Rick. "Folk Models of Home Computer Security." *Symposium on Usable Privacy and Security* (2010), Redmond, WA, 14-16.

[37] Neuhaus, Stephan and Plattner, Bernard. "Software Security Economics: Theory, in Practice." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[38] Wood, Dallas and Rowe, Brent. "Assessing Home Internet Users' Demand for Security: Will They Pay ISPs?" Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[39] Kwon, Juhee and Johnson, M. Eric. "An Organizational Learning Perspective on Proactive vs. Reactive Investment in Information Security." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[40] Gaynor, Martin, et al. "Is Patient Data Better Protected in Competitive Healthcare Markets?" Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[41] Appari, Ajit, et al. "HIPAA Compliance: An Examination of Institutional and Market Forces." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[42] Kwon, Juhee and Johnson, M. Eric. "Security Resources, Capabilities and Cultural Values: Links To Security Performance And Compliance" Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[43] Ashgari, Hadi. "Security Economics in the HTTPS Value Chain." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[44] Wang, Qiu-Hong and Kim, Seung-Hyyn. "Cyber Attacks: Cross-Country Interdependence and Enforcement." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[45] Hunker, Jeffrey, and Probst, Christian. "The Risk of Risk Analysis And its relation to the Economics of Insider Threats." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[46] Stone-Gross, Brett, et al. "The Underground Economy of Fake Antivirus Software." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[47] Baldwin et al. "Contagion In Cybersecurity Attacks." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[48] Florencio, Dinei and Herley, Cormac. "Where do all the Attacks Go?" Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

[49] Herley, Cormac. "The Plight of the Targeted Attacker in a World of Scale." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[50] Florencio, Dinei and Herley, Cormac. "Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[51] Segura, Vicente and Lahuerta, Javier. "Modeling the economic incentives of DDoS attacks: femtocell case study." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[52] Acquisti, Alessandro, et al. "The Impact of Relative Standards on the Propensity to Disclose." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[53] Preibusch, Sören, et al. "The value of privacy in Web search." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[54] Preibusc, Sören, et al. "The privacy economics of voluntary over-disclosure in Web forms." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[55] Malheiros, Miguel, et al. "Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[56] Brandimarte, Laura, et al. "Misplaced Confidences: Privacy and the Control Paradox." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[57] Egelman, Serge, et al. "Choice Architecture and Smartphone Privacy: There's A Price for That." Eleventh Workshop on the Economics of Information Security. Berlin, Germany. 25-26 June 2012.

[58] Egelman, Serge, et al. "An empirical study on user tolerance of security delays." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[59] Tang, Qian, et al. "Improving Internet Security Through Social Information and Social Comparison: A Field Quasi-Experiment." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[60] Bolot, J. C. and Lelarge, M. "A new perspective on internet security using insurance." *Proceedings of IEEE INFOCOM* (2008), 1948-1956.

[61] Böhme, Rainer and Kataria, Gaurav. "Models and measures for correlation in cyber-insurance." Fifth Workshop on the Economics of Information Security. University of Cambridge, UK. 26-28 June 2006.

[62] Shetty, Nikhil, et al. "Competitive Cyber-Insurance and Internet Security." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[63] Böhme, Rainer and Schwartz, Galina. "Modeling Cyber-Insurance: Towards A Unifying Framework." Ninth Workshop on the Economics of Information Security. Harvard University, USA. 7-8 June 2010.

[64] Innerhofer-Oberperfler, Frank, and Ruth Breu. "Potential Rating Indicators for Cyberinsurance: An Exploratory Qualitative Study." Eighth Workshop on the Economics of Information Security. University College London, England. 24-25 June 2009.

[65] Fryer, Huw et al. "On the Viability of Using Liability to Incentivise Internet Security." Twelfth Workshop on the Economics of Information Security. Georgetown University, Washington, D.C. 11 June 2013.

[66] August, Terrence and Tunca, Tunay. "Who Should be Responsible for Software Security? A Comparative Analysis of Liability Policies in Network Environments." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011.

SRI International

Literature Review:
Application of Behavioral Research in Public Policy

April 30, 2014

Prepared for:
Dr. Joseph Kielman
Cybersecurity Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Bincy Ninan-Moses, Roland Stephen, Ph.D., Lucien Randazzese, Ph.D., and Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development
David Balenson, Ulf Lindqvist, Ph.D., and Zachary Tudor
Computer Science Laboratory

Primary Contact:
Lucien Randazzese, Ph.D.
SRI International
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
lucien.randazzese@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

1 Introduction

This literature review represents the second part of a two-part survey conducted by SRI International for the Department of Homeland Security (DHS) Science and Technology (S&T) Directorate Cybersecurity Division (CSD). The first part of the survey, "Literature Review: Current Research in Cybereconomics,"[1] presented to DHS in January of 2014, assessed the extant research in cybereconomic incentives (CEI). CEI research considers cybersecurity activities – good and bad – both through the lens of traditional economics that assumes rational cost-benefit decision making, as well from empirical assessments of actual individual and organizational behavior.

This document takes as its starting point the behavioral research in cybereconomic incentives, and considers a wide range of behavioral sciences and their application to real world public policy issues outside of cybersecurity. Each of the behavior research applications considered in this review involve circumstances, policy challenges, or policy goals with parallels to the circumstances, challenges, and goals associated with improving cybersecurity, particularly the incentives associated with cybersecurity, and thus may provide insight into improving the security and resilience of the nation's critical infrastructure.

The importance of protecting this critical infrastructure was highlighted in February of 2013 when the President issued Executive Order (EO) 13636 on *Improving Critical Infrastructure Cybersecurity*,[2] and on the same day Presidential Policy Directive (PPD) 21 on *Critical Infrastructure Security and Resilience*.[3] These policy documents gave DHS a coordinating role in pursing the cybersecurity objectives outlined in the EO and PPD, and directed the National Institute of Standards and Technology (NIST) to work with stakeholders to develop a voluntary framework for reducing cyber risks to critical infrastructure. They also requested that the Departments of Homeland Security, Commerce, and Treasury identify potential incentives for infrastructure owners and operators to adopt the NIST framework.

The input provided by these departments, as is discussed further in Section 2 below, conforms to a conventional microeconomic view on incentives, one based on rational cost-benefit analysis. While many cybersecurity-related decision making is in fact based on such microeconomic considerations, we know that a great deal more influences the choices people and organizations make beyond what they perceive as the narrow microeconomic considerations of those choices. Examples abound in real life of this "bounded rationality" by which people and groups choose

[1] SRI International, "Literature Review: Current Research in Cybereconomics," January 3, 2014.

[2] The White House, *Improving Critical Infrastructure Cybersecurity*, Executive Order 13636, February 12, 2013.

[3] The White House, *Critical Infrastructure Security and Resilience*, Presidential Policy Directive 21, February 12, 2013.

without perfect (or sometimes any) regard for the costs they will incur or the payoffs they will receive. These examples make a strong case for a more holistic approach to considering the incentives that affect cybersecurity-related decision making and behavior. Taken together, the initial cybereconomic incentives review, in tandem with this document focused on applications of behavioral research outside of cybersecurity, provide this holistic view.

Together these two documents can help enhance the government's ability to address two of the principal challenges in securing critical infrastructure:

1. Ensuring that the individuals and organizations that build, deploy, use, and defend these critical assets are incented to make the best decisions with respect to their security.
2. Creating disincentives to attack the nation's infrastructure on the part of malicious entities that might desire to do so.

This Introduction section is followed by four additional sections as follows:

- Section 2: Choice in Cybersecurity
- Section 3: A Brief History and Overview of Behavioral Economics
- Section 4: Applying Behavioral Sciences in Public Policy
- Section 5: Lessons for Cybersecurity

Section 2 briefly discusses the DHS Integrated Task Force view of incentives that could bear on adoption of the NIST cybersecurity framework, the assumptions made about the nature of decision making reflected in this traditional microeconomic view, and the potential limitations to this view. It then goes on to describe how a broader behavioral view of incentives differs from this traditional view.

Section 3 describes the development of relevant behavioral sciences as a field, and highlights its recent ascendency in influencing several governments in a broad range of policy issues. The scholars that contribute to this area of research tend to be psychologists, sociologists, neuroscientists, and other scientists outside the field of economics per se. Despite this, the field of science that deals with how cognitive, social, and emotional factors affect human decision making is often referred to as behavioral economics. We use this term in this review, but will also refer more generally to behavioral sciences and behavior research when discussing the insights from this research and their applications to policy.

Section 4 discusses several specific insights from the behavioral sciences, as well as numerous examples of their application in specific policy areas. The insights have been organized into six categories which make up the subsections of Section 4. Each of the six subsections briefly characterizes the nature of the behavioral tendencies discussed in each subsection, describes

some of the representative research in each area, and provides a number of examples of how the insights of this research have been applied in the real world. The six categories were chosen as the most effective way to organize what is a large number of specific insights into human cognition, many of which are related to one another. A number of similar ideas, such as Anchoring and Priming in Section 4.3, are sufficiently similar that they have been grouped together in the same subcategory.

Some of these real world applications of the insights of behavioral science described below are quite novel, but many are actually very straight forward – some may even appear obvious. What makes them noteworthy is their level of effectiveness in achieving the behavior-changing goals they set out to address. The absolute number of examples is also very large and growing rapidly. The examples presented do not therefore compose a comprehensive list. Rather, they are representative illustrations of how behavioral science insights are being applied to real world problems, and many come from governments and organizations at the forefront of exploiting behavioral research.

The topics covered in Section 4 overlap somewhat with one another. For example, one of the ways people anchor their cognitive approach to something (Section 4.3) is through the influence of peers (Section 4.6). Similarly, one way in which inadequate self-control (Section 4.5) has been addressed is through various mechanisms that appeal to people's desire to avoid loss (Section 4.2).

In Section 5, at the end of the document, the potential lessons from behavioral science for cybersecurity are briefly considered.

2 Choice in Cybersecurity

Pursuant to its lead role in implementing EO 13636 and PPD-21, DHS established an Integrated Task Force (ITF) to help coordinate interagency, and public and private sector efforts associated with the two Presidential policy documents. The ITF included a working group focused on the topic of incentives. This group, working in conjunction with the Departments of Treasury and Commerce, developed a decision framework (what they refer to as a microeconomic model) designed to map the likelihood that asset owners and operators will adopt the cybersecurity framework under development by NIST. The incentives framework adopts—with qualifications—a standard microeconomic approach. For this reason it offers a good starting point for comparing the usefulness of "neo-classical" assumptions about economic choice with other behavioral approaches in the area of cybersecurity.

DHS defines an incentive as a cost or benefit that motivates a decision. [1] The benefit at the margin of a choice is weighed against the cost at the margin, and each incentive is judged according to how it would impact these marginal choices, as illustrated in Figure 1 below.[2]

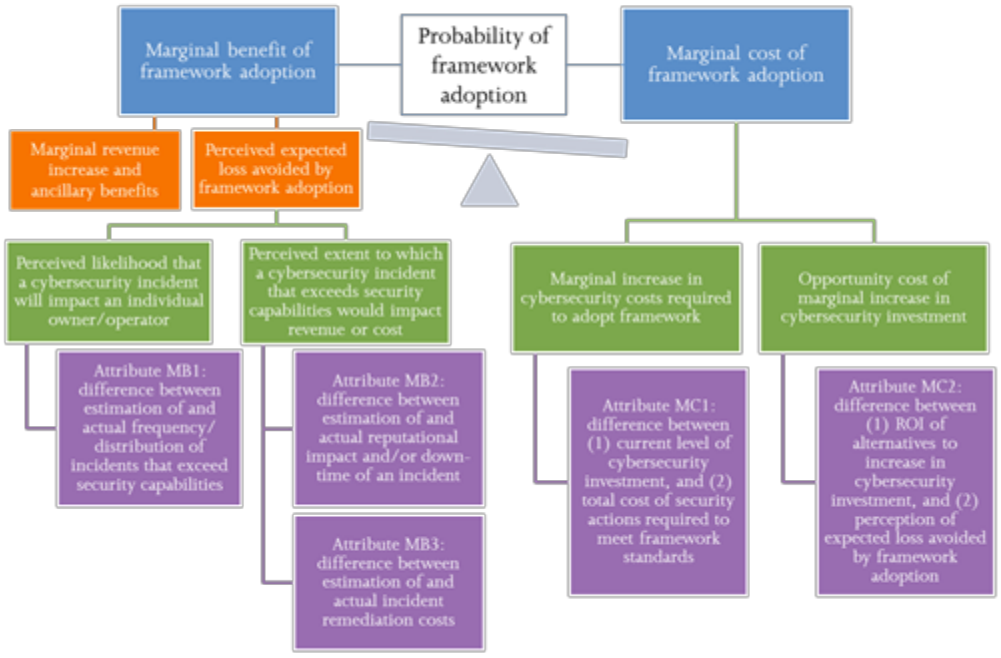


Figure 1: DHS Integrated Task Force Decision Framework

Against the backdrop of this decision framework, the incentives working group considered the impact of ten incentives the government could potentially use to affect the decisions made in the context of the decision framework of Figure 1. These ten incentives are:

1. Grants
2. Rate-Recovery for Price-Regulated Industries
3. Bundled Insurance Requirements, Liability Protections, and Legal Benefits
4. Prioritized Technical Assistance
5. Procurement Considerations
6. Public Recognition
7. Security Disclosure
8. Streamline Information Security Regulations
9. Subsidies
10. Tax Incentives

Conventional “marginalist” frameworks such as that of Figure 1 rely on known costs and benefits. When there is good information, this approach is a powerful starting point for

understanding behavior. Consumers, workers, business leaders, and even criminals are relied on to gather information, calculate their highest payoff, and choose accordingly. That is how markets – legal or otherwise – work. However, Figure 1 highlights perceived as opposed to actual costs. This is inevitable given the lack of data on the scale of loss and the likelihood of loss in the area of cybersecurity. For example, the benefit of adopting the NIST cybersecurity framework accrues, in part, from the "perceived expected loss avoided." What thought processes shape that perception? Assumptions about these processes will determine any estimate of the probabilities of adoption of the cybersecurity framework.

Given the significant uncertainties present in the cybersecurity policy domain, the conventional (or "neo-classical") approach highlighted by Figure 1, and the attendant list of potential government incentives, are of only limited use as a guide to choice, for two related but distinct reasons. Perceived costs reflect uncertainty, and are often formulated where information is missing. This has the effect of generally lowering the value attached to future events, and will make it less likely that an actor will adopt any cybersecurity measure. More than that, however, perceived costs may reflect other influences on behavior, influences allowed a greater role in determining choice when there is little reliable information. For example, an actor may rely on cues from others about what choice to make, or may make choices based on past behavior rather than present circumstances.

These examples of how real world decisions might be made – in the context of peer influence or past experience – highlight still another limitation of the framework articulated in Figure 1, namely that it is focused entirely on incentives at the organizational level. This makes sense given that the framework of Figure 1 was developed explicitly for the purposes of understanding owner-operator incentives for adopting the NIST Framework. But all actual decisions, even when made on behalf of a large government or corporate organizations, are ultimately made by individuals who are subject to a range of cognitive influences outside of those implied in Figure 1.

The usefulness of attention to such behavioral considerations in the domain of cybersecurity extends far beyond the federal cybersecurity framework developed by NIST. As the review of behavioral economics outlined below indicates, information problems and other sources of market failure (externalities and agency problems in particular) mean that informed choice that takes account of all costs is hard, even for sophisticated enterprises, let alone for individual employees and consumers.

2.1 Two Approaches Compared

Table 1 compares, in a stylized way, the two approaches to understanding behavior – microeconomic and behavioral, and is organized according to the same six behavioral ideas around which Section 4 is organized. Each of these six behavioral drivers is described in terms

of the “neo-classical” model and in terms of psychology and social behavior. Both are generally accepted approaches with useful applications, and both can be applied to the domain of cybersecurity (as well as other policy domains). The existing literature on cybereconomics has begun to appreciate the importance of the behavioral perspective (see “Literature Review: Current Research in Cybereconomics”), though much more in this area can be done. The discussion of Section 4 is designed in part to provide the cybereconomic incentives research community with inspiration from what has been done in policy areas outside of cybersecurity.

The key differences between the two approaches described in Table 1 are in the assumptions each makes about the basis of decision making. The “neo-classical” approach assumes agents are well informed (or capable of gathering all necessary information), are capable of calculating payoffs in an unbiased way, and are capable of using the results of their calculations as the basis for rational choice.

In contrast, the behavioral approach assumes that humans have enduring tendencies and biases that shape actions in ways that do not fit a “rational” choice. In situations plagued by information problems, human predilections and biases may play a role that is just as important as rational calculation. Indeed, empirical data clearly show that in many situations human cognitive biases overwhelm our analytical faculties. Thus, in the behavioral view, standard calculations of costs and benefits exist alongside, and may be swamped by, various behavioral biases. Table 1 outlines the differences in assumption between the two approaches.

Decision Factor	Standard (Neo-classical) Microeconomics	Behavioral Science
Choice	People make choices by weighing the cost and benefits of the options they face, and select the optimal choice accordingly	The process of choosing is difficult, and people will often make choices in a way that minimizes the effort in making the choice, with little or no consideration of the actual options
Loss	People treat the prospect of a loss the same as the prospect of a gain when of equal value	Humans are risk averse when faced with a loss and risk accepting when faced with a gain when each are of equal value
Anchoring	People make assessments independent of context	Our evaluations and behaviors are more affected by recent information, experience, or stimuli
Representativeness	People are able to correctly evaluate the impact of randomness	People draw incorrect conclusions about causation and distribution when evaluating random data
Control	People are in full control of the decisions they make and will always behave in a manner they believe to be in their self interest	People will do things they "do not want to do"
Peer Influence	People choose based on the payoffs they will receive, independent of the choices of others	People are susceptible to peer pressure and rely on peers as sources of low-cost information about how to choose

Table 1: Comparison of Classical Microeconomic and Behavioral Approaches

The standard microeconomic approach is very useful. People do often calculate the gains and losses that may arise from one choice or another. That said, there are limits to any individual's ability to reason about the world (so-called "bounded rationality"), especially where information is lacking. As a result, there are real-world examples of people who seem to make decisions without perfect regard for the payoffs they will receive. Behavioral theory is designed to describe and, where possible, explain such actions. What follows is a brief history of this approach, followed by a more detailed review of several of its specific findings and how these have been applied across a variety of public policy domains.

3 A Brief History and Overview of Behavioral Economics

Over the last several decades there has been increasing interest in the application of psychology, sociology, and neuroscience to the understanding of human decision making. This interest has led to a new form of applied research often referred to as *Behavioral Economics*. We can trace the origins of behavioral economics to Nobel Prize winner Herbert Simon's (1955[3] and 1956[4]) seminal research in which he argues what now seems obvious, namely that humans display cognitive limitations and do not have unlimited knowledge or information-processing

capabilities. Simon coined the term "bounded rationality" to describe the limited human capability to solve problems in complex choice situations. He explains that neither do humans have the ability to access all available information in complex choice situations nor do they possess the processing power required to assimilate all available information to make a choice that results in maximum benefits.

Simon shows that in complex choice situations humans can respond well enough to what he calls "satisfice" – combining the words "satisfy" and "suffice"; but not well enough to truly "optimize" or select the best possible option. Using the understanding of bounded rationality to examine the capability of humans is a more realistic approach to understanding human behavior and decision making. It stands in contrast to traditional microeconomic theory which suggests that "economic man" will always make the best decision that renders maximum benefits when presented with options, even in complex choice situations.

Research in behavioral economics gained further momentum following the works of psychologists and behavioral scientists, Daniel Kahneman and Amos Tversky[5] (1974) on economic decision-making. Daniel Kahneman, known as the father of behavioral economics, won the Nobel Prize in Economics in 2002 for pioneering work in this field. In his book *Think Fast, Think Slow*[6] (Kahneman 2011), Kahneman identifies two types of thinking in humans:

- ***System 1* or *automatic thinking*** which portrays our irrational, fast, and intuitive behaviors in decision making; and
- ***System 2* or *reflective thinking*** which represents our rational, slow, and derivate form of decision making

Most of the time, we rely on automatic thinking (system 1) to execute routine tasks such as navigating our daily morning commute or communicating in our mother tongue. Our reflective mind, on the other hand, is responsible for the decisions into which we put more conscious thought (system 2), like whether or not to go to the gym or how to respond to a child's request to stay out late.

Economist Richard H. Thaler and legal scholar Cass R. Sunstein helped to bring behavioral sciences into mainstream, non-academic thinking, applying its insights to policymaking in a number of countries as well in the private sector. Thaler and Sunstein introduced the concept of "choice architecture," which describes how the decisions people make can be influenced by how the choices or options themselves are presented. They also introduced the idea of "nudging," or the intentional steering of people towards a particular decision in a way that does not actually change the explicit incentives associated with any of the options. For example, a range of snack options, including both health and unhealthy ones, can be provided to a group of people. Making the fresh fruit the easiest to reach on the table provides a nudge to people to select this

healthy option, but in no way changes the value of the options or the incentives associated with selecting among them.

Choice architecture makes the assumption that people are irrational, imperfect, and susceptible to influences from their environment, and uses that understanding to influence the choices people make. Thaler and Sunstein's concepts of choice architecture and nudging piqued the interest of a number of policymakers and the private sector managers, and were instrumental in inspiring the formation of "nudge" units in countries like United Kingdom (UK)[7], France[8], Denmark[9], and United States[10] to understand how these concepts could be used to affect better decision making among their citizens.

Highlighting the growing recognition of how behavioral factors can be effective in the design of policy interventions, in 2010, the UK established the Behavioral Insights Team (BIT) within the Cabinet Office of the Prime Minister. The goal of BIT was to "Test, Learn, Adapt" behaviors in pursuit of various policy objectives.[11] BIT collaborates with various public and private sector organizations and uses randomized controlled trials to test and evaluate behavioral science inspired public policy interventions in the areas of public health, public finance, energy efficiency, education, consumer spending, and retirement savings. BIT evaluators use trials to understand the impact of interventions and refine these interventions prior to recommending actual policy changes.

In 2006, the French Prime Minister setup a similar behavioral science focused unit called the Center for Strategic Analysis, now known as the General Commission for Strategy and Economic Foresight, to understand how behavioral sciences can inform public policy. In 2013, the U.S. White House Office of Science and Technology Policy followed suit and started hiring social scientists to work with the US government to *strengthen federal capacity for behavioral insights*.[12] Many other counties, such as Denmark, Netherlands, and Singapore, have also been developing policy interventions using lessons from behavioral economics.

In March 2014, Martin Wheatley, Chief Executive of the UK Financial Conduct Authority, urged the financial services industry to make greater use of the insights from behavioral economics to understand what drives client decision making in financial services.[13]

4 Applying Behavioral Sciences in Public Policy

This section discusses several of the most important, widely studied, and widely applied ideas that have come from behavior research, and is organized around the specific concepts outlined in Table 1. In each of the six subsections below, a brief overview and definition of the concept under consideration is provided and its applications in policy are described. In a number of cases, applications have been done in partnership with organizations in the private sector.

4.1 Choice Fatigue and Defaults

Researchers refer to the difficulty people have in making decisions in the face of many competing options as *choice fatigue*. Choice fatigue often leads to a deteriorating quality of decision making by an individual after a sustained period of decision making, and has been shown to occur in a wide range of specific settings. On experiencing choice fatigue, people tend to select options that require the least effort to be exerted in the decision making process. In other words, decisions are based on what makes the act of choosing itself as easy as possible, not on the value of the various options. In such situations, researchers find that people prefer so-called "default" options, which can be described as options that have already been selected for people by the option provider (Thaler and Sunstein, 2008[14]).

In an interesting illustration of the choice fatigue, Iyengar et al.[15] (2004) studied the impact of Vanguard's provision of additional fund choices in 401(k) plans by analyzing 800,000 employees across companies whose 401(k) plans were administered by Vanguard. The authors found that with every 10 additional funds in a plan, there was a decrease in employee plan participation by 1.5-2 percentage points. When employees were faced with a large number of options, they found it difficult to choose and declined to participate.

Levav et al.[16] (2010) conducted a field experiment in which customers had to pick from 67 different car attributes such as interior fabric color, exterior color, sound and navigation system options, types of engine and so on, on a computer before purchasing a vehicle. Each option had a default option preselected. As people progressed through their choices they started to select the default option more and more frequently. The authors show that this choice fatigue occurs not only when making a series of directly related decisions, but that people exhibit choice fatigue for an unrelated set of decisions if they've faced a significant amount of decision making in a completely different domain immediately beforehand. These effects can endure for long periods of time, so that people who tend to face a lot of decision making in their daily may face choice fatigue even for one-off decisions.

Danzinger et al.[17] (2011) have shown how human decision making capability declines over time in the case of parole hearings. The authors show that the likelihood of ruling in favor of a prisoner declines as the day progresses. However, after the court lunch break, parole approval rates increase, but then again start to decline over the course of the afternoon. This tendency towards default option selection has been highlighted by other researchers as well (cf. Park et al. 2000[18], Johnson & Goldstein 2003[19]).

As described in the previous section, Thaler and Sunstein[20] (2008) define choice architecture (explained in the history section above) as a way for policymakers to structure choices based on lessons from behavioral sciences to influence the choices people make. Choice architecture can be something as simple as the placement of options on a list. For example, Meredith and

Salant[21] (2011) demonstrated that the candidates listed first on a ballot are most likely to win an election.

In experiments conducted in the U.S., researchers have shown that decisions involving everything from organ donation (Johnston & Goldstein 2003[22]), to 401(k) participation and associated asset allocation (Madrian & Shea 2001[23], Choi et al. 2004[24]) can be significantly influenced through the use of defaults. In each these cases, people were enrolled in the programs in question (organ donation and 401(k) savings plans) by default, resulting in correspondingly higher organ donor rates and savings rates. People were also free to opt-out of default enrolment if they preferred not to participate, but most people did not make the effort to opt-out.

UK's BIT used insight from studies such as those described in the preceding paragraph to change the National Health Service's organ donation program to an "opt-out" system by which donors were enrolled as organ donors by default. Traditionally, organ donor registration rates were very low in UK despite survey data showing that 65% of the population was prepared to enroll as organ donors. The new opt-out system increased donor registrations by 20% in 2008.[25] The states of Illinois, Texas, and California in the US similarly switched to opt-out donor systems. People are again free to opt-out if they prefer not to participate. These states were, explicitly counting on choice fatigue, and across these states organ donor registrations increased by more than 50%.[26]

Bertrand et al.[27] (2010) conducted a field experiment in South Africa that provided various lending and payment structure options for consumer loans. The authors found that providing just a single loan option instead of four different options increased participants' decision to apply for a loan by as much as a 2 percentage point reduction in interest rates.

Exploiting the concept of choice fatigue and the propensity of people to select default options, Thaler & Benartzi[28] (2004) designed the Save More Tomorrow (SMaRT) program in which employees were, by default, signed up to participate in retirement savings and to increase their retirement savings rate whenever they received a pay raise, up to a preset limit. Participants were free to opt out of this program if they preferred not to enroll. The SMaRT program increased annual savings by an estimated $7.4billion. In a similar program in the UK, 90% of employees who were automatically enrolled in a pension scheme, where the choice to opt-out was also provided to them, continued to participate in the scheme after their automatic enrollment.[29]

Bettinger et al.[30] (2011) conducted an experiment to understand how choice fatigue and the strategic use of defaults can impact applications for college financial aid, and by implication, rates of college enrollment. The authors hypothesized that filling Free Application for Federal

Student Aid (FAFSA) forms is intimidating to many low to moderate income individuals because of the threat of federal penalties of up to $20,000 for inaccurate information. During the research, participants were provided personal assistance to pre-populate FAFSA forms by tax professionals as part of the personal tax preparation process. Following this assistance with pre-populating the forms, college enrollment of low to moderate income individuals increased by 29% for two consecutive years.[31]

In a program reminiscent of the hardware-enabled security concept, the UK's BIT recommended changing the defaults for heating and cooling systems used in government buildings. New systems are designed to stay on only during regular work hours when the buildings are open for business. Employees have the option to keep the systems on if they have to stay beyond normal business hours. Using these defaults in heating and cooling, government departments in UK reduced carbon emissions by 10%[32].

4.2 Prospect Theory and Loss Aversion

Loss aversion is the human tendency to outweigh losses relative to equally sized gains. In traditional utility theory, under circumstances of risk or uncertainty, people are assumed to rationally weigh the outcomes of different options to make the best decision. Behavioral scholars Kahneman & Tversky (1971[33] and 1979[34]) introduced an alternative model called Prospect Theory. In their research, the authors show that when faced with risky or uncertain situations, people tend to underweight potential gains or rewards and over-value potential losses. Specifically, people make decisions based on the potential value of losses and gains rather than final outcomes. Thus people may prefer a scenario that is identical to another when that scenario is described as getting to some outcome via a gain than arriving at the same outcome in terms of a loss.

We also see this discrepancy in value in Locke and Mann's[35] (2003) research, in which the authors discuss the phenomenon in the context of financial markets. The authors analyzed trades made by 334 professional traders and found that the propensity of these traders to hold on to falling shares was longer than rising ones. The authors explain this behavior in terms of the traders' desire to avoid losses, and go on to show that this tendency lowers the long-term financial performance of the portfolios these traders manage.

The impact of loss aversion is highlighted by the efforts of the credit card industry to change pricing practices, first through legislation and then through systematic marketing efforts. In 2008, the credit card lobby in the US tried to get a bill passed in Congress prohibiting charging fees for credit card transactions. The credit card lobby argued that it was detrimental to the credit card business to penalize customers for credit card purchases, but was unsuccessful in its lobbying efforts. As Thaler[36] (1980) explains, the industry understood that people are more average to realized losses than realized gains, so embarked on a systematic effort to flip the popular

perception of and nomenclature for the added fees many business change for credit card payments on their heads. It promulgated the idea that this price delta was due to a discount for using cash. People are less affected by the loss of a gain (cash discount) than they are a realized loss (fee for using a credit card). The industry believes, correctly, that the extent to which it can get consumers to think in terms of forgone cash discounts, they will be more likely to use a credit card.

In Philippines, Committed Action to Reduce and End Smoking (CARES) is a savings program offered to smokers by the Green Bank of Caraga that exploits the loss aversion principle to reduce smoking. People who enroll in the program open an account with a minimum balance of a dollar and are required to add money to the account every month for six months. This monthly deposit amount reflects the amount that the smokers would otherwise pay in a month to purchase cigarettes. At the end of the six month period, each account holder has to take a test to confirm that he or she has not smoked. If he or she fails the test, the person loses the money to charity; otherwise, the amount is returned in full back to the person[37].

Loss aversion has also been used to explain poor adoption of energy saving technology by consumers, and the UK's BIT used this behavioral understanding to improve adoption. Research by the BIT indicates that people are not installing energy efficient products because doing so involved significant investment up front, whereas the payoff for this investment takes an extended period of time[38]. Even if one considers the discounting of future gains, consumers fail to make energy saving investment even when the future benefits would significantly outweigh the upfront costs. The reason is that people are not merely discounting future benefits, but they are cognitively inflating the true cost of the upfront cost by thinking of them as a loss. The UK government's Green Deal program provided people with an interesting option to install energy efficient products in their homes at no upfront costs. The Green Deal gives people the option to pay for the changes in small portions over time in their energy bills, and significantly improved adoption rates.

In the case of cybersecurity in the workplace and loss aversion, Herley[39] (2009) argues that people generally do not take the effort to practice secure behavior online because they do not see immediate gains. Instead people find that they lose out on productivity when they take additional time to make sure that their online behavior complies with corporate security practices.

4.3 Anchoring

Anchoring refers to the tendency to rely heavily on recent experience, stimuli or available information, even if incorrect or limited, to make decisions. As with much of the research in behavioral decision making, Tversky and Kahneman[40] (1981) were among the first scholars to extensively study anchoring. As will be seen in the examples below, in some situations anchors

are established unconsciously, and in other cases, especially those involving uncertainty, people use available and salient anchors from their environment to help make decisions.

A study of anchoring by Nunes & Boatwright[41] (2004) provides a good example of how the anchoring phenomenon manifests in people regardless of education levels. The researchers gave MBA students a bottle of wine and asked them to value it. But first, the students were asked to recall the last two digits of their social security number. On average, the students, whose last two digits of the social security number fell below 50, were willing to pay $11.62 for the bottle of wine and the students whose last two digits of the social security number fell above 50 were willing to pay $19.95. In this situation, participants' acknowledgement of the last two digits of their social security number provided an unconscious anchor that affected their subsequent estimations.

In a study that highlights how anchors arise when people consciously look for apparently relevant information, Thaler & Sunstein[42] (2008) asked students at the University of Chicago to estimate the population of Milwaukee, Wisconsin. The authors noted the analysis process that the students used to arrive at the population estimate. Some students relied on their knowledge of the population of Chicago, which was around three million. These students considered the size of Milwaukee: though Milwaukee is smaller than Chicago, it is the largest city in Wisconsin. Students arrived at roughly one million after estimating that Milwaukee's population is roughly one-third of Chicago's population. Another set of students who were from Green Bay, Wisconsin approached the question differently. They started the process by considering the population of Green Bay, which at the time of the experiment was roughly hundred thousand. They then estimated the population of Milwaukee to be three times that of Green Bay, and arrived at roughly three hundred thousand. The two sets of students started the process of estimation from low or high anchors using unique available knowledge and so arrived at widely different estimates. The actual population of Milwaukee at the time of the experiment was 580,000.

Englich & Mussweiler[43] (2001) found that even experts can be prone to anchoring. To study this, they gave two sets of experienced judges the same written account of a rape case. One set of judges was told that the prosecutor demanded a sentence of 12 months, and the second set of judges was told that the prosecutor demanded a sentence of 34 months. Both sets of judges were asked to render an opinion on sentencing. The first set of judges awarded sentences on average of 28 months and the second, an average sentence of 36 months.

Closely related to the concept of anchoring is that of priming. Priming refers to the way in which our behavior with respect to one stimulus can be influenced by earlier exposure to another stimulus. For example, when researchers exposed selected participants in an experiment to references of old age, through words and images, and then ask participants to walk down a hallway, researchers find that those exposed to the "aged stimulus" walked more slowly, or

more like older people, that those participants who were not exposed to any references to old age. Some scholars describe priming as the underlying mechanism by which anchoring occurs (cf. Mussweiler & Strack,[44] 1999).

A number of researchers have shown that people are more likely to do something when provided with an associated priming stimulus. Sherman[45] (1980) has shown that people can be primed to be more likely to act in a certain way when they are questioned about their intentions. The author found that merely asking people to predict their compliance with socially desirable behavior increases their likelihood of subsequently performing the behavior in the future. For example, Sherman asked people whether they would donate time to charity in the future, and found that people were more likely to do so if asked the priming question about their intentions in this regard. Greenwald et al.[46] (1987) found that campaign officials were able to increase the probability of voting by 25% if they asked people whether they intended to vote a day prior to election. Berger et al.[47] (2008) found that when voting was conducted in actual schools (versus some other voting location), voters were more likely to vote in favor of a ballot initiative to raise taxes for purposes of increased expenditure on education.

As indicated in the introduction, many of the real-world application of behavioral science insights are quite simple in design, yet also quite effective despite this simplicity. For example, one of the earliest experiments into priming was conducted by Leventhal et al.[48] (1965). The researchers examined awareness among Yale seniors of the risks of tetanus and the value of inoculation. After an information session on the importance of inoculation, students were asked if they intended to get inoculated. While many said they did, only 3% acted on their stated intention. In a subsequent information session, a second set of students were also given a copy of a campus map with the location of the health center highlighted on the map and again asked if they intended to get the tetanus shot. Priming the students with the map and location of the health center resulted in 28% of the students getting the tetanus shot, a 930% increase in the rate of inoculation. The only difference between the two scenarios was a map showing the location of a health center, whose location was already known by all the students.

Priming has been shown to be impactful in influencing virtually any behavior. It has long been known, for example, that people are more likely to purchase insurance after a recent natural disaster (Slovic, Kunreuther, White[49] 1974). Less expected, however, have been the results of research that shows that people are less likely to cheat, more likely to forgive, and more likely to donate to charity after exposure to religious primes such praying or reading from a holy book (Shariff and Norenzayan[50] 2007, Randolph-Seng & Nielsen[51] 2007, Pichon et al.[52] 2007).

In a powerful example of how priming can be used to encourage desired behavior, Payne[53] (2010) used a simple line of yellow tape along with a small sign to allocate part of supermarket shopping carts for fruits and vegetables. The visual prime of the separate location in the cart

designated for fruits and vegetables actually resulted in shoppers purchasing more fruits and vegetables. Consistent with Payne's experiment, the BIT is partnering with UK supermarkets and other retailers to promote healthy purchasing among customers through the use of shopping cart segregation and other similar primes[54].

In Sweden, Volkswagen made use of priming stimuli to encourage metro passengers to use the stairs rather than elevators or escalators. Volkswagen designed the stairs in the Stockholm metro station using motion-sensor piano keys that played musical tones when people climbed them[55]. Volkswagen also painted the stairs to look like piano keys. Peeters et al.[56](2013) found that using fun visual and audio components, Volkswagen inspired 66% more passengers to use the stairs.

This type of emotion based priming has been used to change behavior with much more serious consequences. Via a program called Babies of Borough in Southeast London, the faces of babies from the neighborhood were painted on building shutters in the neighborhood, resulting in a reduction in looting by 20%. OgilvyChange, the behavioral economics arm of the advertising firm, Ogilvy & Mather, was responsible for exploiting the use of priming in the campaign.[57]

4.4 Representativeness

Representativeness refers to the heuristic by which people tend to determine the probability or frequency of an event based on general assumptions or past experience. This heuristic comes into play frequently when the quantity that people are attempting to assess is influenced by randomness. Thaler & Sunstein's[58] (2008) work with regional cancer clusters highlights this situation. It is often the case that particular neighborhoods report more cases of cancer in a given period of time than adjacent neighborhoods. This leads people to assume that there is some systematic causal reason for the disease to be prevalent within one small area and not another, such as localized contamination of some sort, and to call for investigations of possible local causes such as contamination. However, random variation alone will always lead to higher incidence in some areas versus others, independent of any physical difference in risk across these areas.

Gilovich[59] (1991) describes a similar example of London residents who lived along the Thames River during World War II. German missiles seemed to have landed the most along the Thames River during the war, and the pattern of the missile hits led to the belief that the Germans could aim their bombs with great precision and that areas where the bombs didn't land were probably pockets where German spies lived. However, statistical analysis confirmed that the strikes were actually consistent with a random distribution. Representativeness leads people to construct causal patterns where none exist, and to make assumptions before looking at evidence.

A classic example of the way representativeness can lead to false reliance on past events occurs in games of chance, and is known as the gambler's fallacy. [60] Gamblers will often assume that after ten consecutive fair coin tosses that result in heads, the probability that eleventh toss will result in heads as well is very low or close to impossible. Of course the odds that the next coin flip results in a head is the same as any other toss, but representativeness leads people to believe past events (a series of heads) somehow affects future events. Though called the gambler's fallacy, the same basic representativeness phenomenon occurs in any setting in which people base future estimations about chance events on unrelated past events. More generally, representativeness highlights the fact that most people are not able or inclined to correctly assess situations in which randomness is a central feature.

An interesting 1995 study that combined elements of both priming and representativeness was conducting focusing on binge drinking on college campuses[61]. General consensus among students was that their peers in college consume a lot of alcohol, as it turns out more than they actually do. The University of Arizona, concerned with binge drinking rates on campus, decided to determine the actual amount of alcohol students consumed. Researchers found that the actual rate was much lower than assumed levels. By merely releasing these data widely and thereby challenging campus assumption about the actual amount of alcohol consumed, the university was able to reduce binge drinking rates on campus by 29%[62]. The Welsh Assembly Government and the charity Drinkaware are working on a similar effort to reduce alcohol consumption among university students at Wales[63].

4.5 The Illusion of Control

People have a well-established propensity to overestimate the control they exercise over situations, a propensity behavioral scientists often refer to as the *illusion of control*. This lack of control often manifests itself as what people commonly think of as lack of self-control (Gino et al. 2011[64]). The lack of control can happen in situations in which no exogenous factors other than self-control affects the outcome, such as when someone tries to quit smoking. The illusion of control can also happen in situations in which we have no control at all, such as when people prefer to pick their own lottery numbers than have others pick for them, in the mistaken belief that they are better at picking "winning" lottery numbers (Dunn and Wilson 1990[65]).

Policymakers try to design commitment mechanisms for self-control issues as a way to get people to commit to decisions they make regarding issues affecting the public interest, often public health challenges like quitting smoking or losing weight. Setting up situations in which people commit "publicly" to action setting has been found to be effective in many settings and counties. The UK's BIT, the pharmacy chain Boots, and the U.K. Department of Health are working together to conduct smoking cessation trials using lessons from the CARES program in the Philippines (mentioned in Section 4.2 on loss aversion). Researchers in the US also

conducted several experiments, using lesson from the CARES program, which were very effective in helping smokers end their addiction (Gine, Karlan, and Zinman 2008[66]). Other mechanisms of public commitment, such as asking smokers to sign, in a public ceremony, a pledge to end their addiction to smoking, have been shown to increase rates of smoking cessation[67,68].

Appreciation for the widespread lack of sufficient self-control has led to a large number of tools being developed to address issues of control. The operating premise underlying these tools is the behavioral science of self-control, and its relative scarcity. As with some of the application examples in the preceding sections, these tools are often very simple in their execution details, but nevertheless have a significant impact on behavior.

Stickk.com is a self-control aiding website started by economists to help people meet goals that are more easily achieved with improved self-control[69,70]. The website provides financial and non-financial commitment mechanisms to help people accomplish specific goals, often within certain time periods. With the financial commitment mechanism, people risk losing money to an organization they dislike or political campaign they don't support if they are unable to achieve their stated goals. With the non-financial commitment mechanism, people risk facing the disapproval of friends and family. If they are unsuccessful in meeting their goal, an email stating as much is sent out to a group of their friends and family.

"Clocky" is typical of a number of technological devices developed to address the lack of control[71]. It's an alarm clock on wheels that literally runs away when the snooze button is used too much. It can be programed for an allowable number of snoozes, and after the set number of snoozes is reached, the clock jumps off the night stand and moves around the room making an annoying sound.

Knowing that many people lack the self-control to stop engaging behaviors they do not want to engage in, several states, including Indiana, Illinois, and Missouri, have created gambler registers on which gamblers voluntarily put themselves. Anyone on a register is refused entry into a casino[72].

4.6 Peer Influence

People are prone to the *influence of peers* in many situations, especially those involving uncertainty. Research shows that people are influenced by the actions of other people, by their preexisting expectations of what other people would do in similar situations, by pressure exerted by peers, and by social norms (Cialdini et al.[73] 2006).

Some of the very earliest behavioral research was in the area of peer influence. As far back as the 1930s, Sherif[74] (1936) conducted a series of experiments to understand how peer influence

affects cognition. In the first experiment, participants were asked to sit in a dark room, focus on a pinpoint of light, and determine the distance that the light moved from the initial point while participants were in the room. Participants' estimates of the distance the point moved varied widely. In the next step of the experiment, participants were asked to join small groups, and then provide their individual estimates of the distance moved in a public setting. The individual estimates provided in the group setting were different from the initial response that each individual participant had provided when they were alone. In fact, the responses had less variation and instead tended toward a group median. In the third step of the experiment, the author introduced a point of influence to each group, an individual who was confident and outspoken in his estimate, but was not actively trying to change other participants' estimates. Sherif found that when the point of influence's estimate was higher than other participants, the group conformed to a higher estimate; and in the other case, when the individual's estimate was lower, the group agreed upon a lower number.

In a similar study of peer influence, Asch[75] (1951) conducted a series of experiments of people's response to questions in various settings. In the first experiment, individual participants were seated in private rooms when asked to respond. In a second experiment, individuals were asked to sit in a group setting and respond to the same questions. When individuals realized that others in the group provided different answers, the individuals tended to change their original answers to be more consistent with those of the group, even though the people participating in the experiments were strangers. The authors argue that even in the case of presence of strangers, people worry about the social disapproval of others in a group. Because of this peer influence, participants were willing to change answers they actually believed were accurate to start with, giving what they thought were less accurate responses in order to conform with the group.

An interesting example of the effects of peer influence has been studied from 1954. During that year there was an apparent epidemic of windshield pitting which turned out to be the effects of peer influence[76]. A relatively small number of people in the Seattle area noticed small holes or pits in their windshield and reported these to local authorities and news outlets. As news of the apparent increase in windshield pitting spread across the state of Washington, people from nearby cities began reporting similar incidents. As the reported incidents became more widespread, people started to assume that the pitting was due to widespread acts of vandalism, radioactivity, an odd atmospheric event, or some other systematic cause. However, once scientists investigated the phenomenon in detail, they found that the windshield pits were present on these cars all along, and were only noticed as reporting became more widespread, leading to further reporting, and so on.

The cognitive biases associated with binge drinking and its mitigation have been studied in detail by a number of scholars (e.g. Straus & Bacon 1953[77], Wechsler et al. 1994[78], and

Wechsler & Nelson 2000[79]), with peer influence associated with both the high incidence of binge drinking and with strategies for its successful reduction.

As a consequence of reported high binge drinking rates on campus, in 1995 the University of Arizona (UA) decided to assess the actual rate of drinking on campus. Researchers at UA conducted a set of surveys to determine actual levels of campus alcohol consumption, and students' perceptions of how much alcohol their peers consumed. These surveys showed that students actually misperceive campus norms and attitudes regarding alcohol consumption, and greatly overestimate the levels of consumption by their peers.[80] UA published this information widely in an attempt to change the social norm regarding binge drinking. As discussed in Section 4.4 above on representativeness, when beliefs about the frequency of binge drinking are changed (i.e. lowered to more accurate levels), peer influence can be used to reduce binge drinking, as the University of Arizona openly did when it published the actual amount that students drank. This sharing of the actual data on alcohol consumed helped to reduce drinking on campus by more than 29%.[81]

Looking not at the influence of strangers but of those who are close, the research of Thaler & Sunstein[82] (2008) highlights how people make financial or health-related decisions based on what their friends and family do. Researchers such as Redhead[83] (2008) suggest that stock market crashes result in large part because of the peer influence investors have on one another. What's interesting about this analysis is that this peer influence happens not only among uniformed investors, but also among market professionals. Redhead describes the scenario by which these professionals begin to reject conventional methods of share evaluation, ultimately creating a bubble which eventually pops. Thaler and Sunstein (2008) have argued for a stronger role for public policy in the face of evidence that people are exhibiting the type of behaviors that cause dramatic movements in markets.

Opower is a software-as-a-service company that works with utility providers around the world to promote energy efficiency among utility customers. In the US, Opower exploits the power of peer influence to reduce energy consumption among the customers of its utility clients[84]. Opower sends detailed energy consumption reports to every utility customer, ranking households against their neighbors' consumption. When people find out that they fall in the above average energy usage group, they reduce future consumption by 2-3%. In contrast, consumers who fell in the below average usage group sometimes used more energy the following month. Accordingly, the program was less effective than it could be, with increased usage on the part of low usage customers offsetting gains from reductions from high usage customers. In order to address this problem, Opower added a sad smile icon to the consumers' rank if the consumer fell in the above average usage group and a happy smile icon for the below average usage group. These icons had the effect of transmitting social approval, approval that customers wished to maintain. This usage of social approval encouraged consumers to

keep their energy usage lower. As a result of this behavioral intervention, 600,000 Opower client customers are using a sustained 2% less energy.

Peer influence has been used in a number of settings to encourage payment of taxes. In Minnesota, informing taxpayers that the actual rate at which people paid their taxes on time was higher than perceived levels helped increase voluntary tax compliance in the state. Similarly, the Irish government encouraged 35.5% of tax-delinquent pub owners to renew their pub licenses merely by sending them letters with information about the compliance rates of their fellow pub owners[85,86]. The UK government loses several billion pounds annually to uncollected taxes. The BIT worked with Her Majesty's Revenue and Customs' Risk and Intelligence Service Campaigns Team to develop a trial directed at medical professionals who owed the government outstanding taxes. BIT sent the sample group of medical professionals personalized letters emphasizing the outstanding tax amount, that tax compliance among doctors was 97%, and that the general public perceived medical professionals to be honest and trustworthy. The experiment led 35% of delinquent doctors to pay their taxes[87].

In February 2014, the state of Maryland published the names of tax defaulters and amounts they owed the state online.[88] The state hopes that because people care about social disapproval, the strategy will shame tax defaulters into paying their taxes. The government in the UK has also published the names of tax evaders on public websites or in trade journals. They have done the same thing with people who have been found guilty of crimes of financial fraud. In each of these cases the aim has been to exploit exposure to social disapproval as a means to exert peer influence on behavior. [4]

A number of organizations have built peer influence components into campaigns designed to encourage people to engage in some desired behavior. In principle each of these programs could have been designed without inclusion of a peer component, but by leaving this component out would have left each much less effective than they turned out to be.

Nike launched the Nike+GPS iPhone app which people can use to monitor their running, and integrated it with a competition called the Grid in 2010, to turn routine exercise into a peer comparison event. As part of the competition, runners are required to scan codes on phone boxes installed in various locations in London as they progressed through the competition. Participant progress in the race to scan each phone box location was published on the Grid

[4] In the context of cybersecurity this type of exposure to social disapproval is sometimes referred to as shaming. The DHS presentation, "DHS Incentives Study: Analysis, Recommendations, and Areas Identified for Further Research," (August 22, 2013) indicates that there is "little evidence of effectiveness independent of procurement requirements and potential for unintended consequences such as cyber targeting."

community website. The success of this program has led to the formation of running clubs around the country[89,90].

Two organizations in London, Transport for London and Intelligent Health, collaborated to create the "Step2Get" initiative that uses swipe card technology, online gaming, and rewards to encourage students to walk to school, and relies heavily on the peer influence effect[91]. Swipe card readers have been installed on lamp posts along the route to schools in the city. Students are encouraged to walk to school and participants are required to swipe their cards against these readers along their way to school for points. As with the Nike running program, students can check their accumulated points online and compare how they are faring against others in their school. The program was so successful that it has led to an international version of the program called "Beat the Street."[92] In "Beat the Street", students from schools in cities around the world, including New York, London, Shanghai, Liverpool, and Southampton, are competing to see who can rack up the most miles to and from school over the year.

Finally, Facebook, being in a natural position to make use of peer influence, did so in a campaign to encourage voting during the 2010 U.S. elections. Around 60 million users received a "social" message that included the profile pictures of six randomly selected Facebook friends who had clicked the "I voted" button that day. Another group of users merely received a message encouraging them to vote, absent any information on what their friends (i.e. peers) had done, and a control group received no message at all. Post-election analysis indicated that receipt of the social message led to an estimated 60,000 additional voters, and a further 280,000 people were indirectly nudged to the polls by seeing messages in their news feeds, for example, messages telling them that their friends had clicked the 'I voted' button.[93]

5 Lessons for Cybersecurity

The concepts described above portray a wide range of situations in which human decision making is influenced by psychological rather than purely objective analytical factors. Policymakers have exploited these behavioral biases to influence behavior in areas as diverse as public health, energy conservation, retirement and financial planning, fraud prevention, tax collection and several others. Equally wide ranging has been the geographic scope of the policy application of behavioral theory; its concepts have been used by governments in North America, Europe, and Asia.

These examples highlight a number of concepts that have potential application in cybersecurity generally, and in particular in managing the incentives that govern cybersecurity behavior and investment. Table 2 below reproduces part of Table 1 above, summarizing, for each of the decision factors covered in the subsections of Section 4, the basic insights from behavioral

science. The rightmost column of Table 2 highlights the implications of each of these insights for cybersecurity incentives.

Some of the cybersecurity applications of the ideas presented in this review may be quite simple – for example, the lessons of choice architecture, choice fatigue, and the power of defaults suggests design principles that make the highest level of security a universal default. Other potential obvious applications might be more difficult in the real world to apply. For example, while shaming has proved effective in getting people to pay their taxes, in the cybersecurity context it raises questions about potential target identification for would-be attackers. In the long run, developing effective behavioral science inspired cybereconomic incentives policies will require combining lessons from behavioral research with deep subject matter expertise in cybersecurity.

Decision Factor	Insight from Behavioral Science	Implications for Cybersecurity Incentives
Choice	The process of choosing is difficult, and people will often make choices in a way that minimizes the effort in making the choice, with little or no consideration of the actual options	Superior security options may not be selected if that selection required too much analytical effort
Loss	Humans are risk averse when faced with a loss and risk accepting when faced with a gain when each are of equal value	People may react more strongly to incentives conveyed as protection from losses than those seen as rewards or payments
Anchoring	Our evaluations and behaviors are more affected by recent information, experience, or stimuli	Measuring and broadly promulgating the nature and consequences of security breaches could help overcome anchoring -based complacency with respect to security
Representativeness	People draw incorrect conclusions about causation and distribution when evaluating random data	Assessment of the location (in time and space) of cybersecurity risk and the impact of this risk may be biased
Control	People will do things they “do not want to do”	Policymakers and organizations should not assume people will avoid behavior merely because a behavior or its consequences are detrimental to people’s self-interest
Peer Influence	People are susceptible to peer pressure and rely on peers as sources of low-cost information about how to choose	Cybersecurity-related incentives that require group-wide compliance or performance may be more effective than incentives aimed at individuals alone

Table 2: Implications of Behavioral Science for Cybersecurity

Research Referenced in this Review

[1] See "DHS Incentives Study: Analysis, Recommendations, and Areas Identified for Further Research" Tony Cheesebrough, National Protection and Programs Directorate and Integrated Task Force for EO 13636 and PPD-21 (August 22, 2013)

[2] Figure 1 drawn from "DHS Incentives Study: Analysis, Recommendations, and Areas Identified for Further Research" Tony Cheesebrough, National Protection and Programs Directorate and Integrated Task Force for EO 13636 and PPD-21 (August 22, 2013)

[3] Simon, Herbert A. (1955): "A Behavioral Model of Rational Choice", Quarterly Journal of Economics 69(1), pp. 99–118.

[4] Simon, Herbert A. (1956): "Rational Choice and the Structure of the Environment", Psychological Review 63(2), pp. 129–138

[5] Tversky, Amos and Kahneman, Daniel (1974): "Judgment under Uncertainty: Heuristics and Biases", Science, New Series, Vol. 185, pp. 1124-1131

[6] Kahneman, Daniel (2011): "Thinking, Fast and Slow", Farrar, Straus and Giroux, 2011

[7] See, The Behavioral Insights Team, UK Cabinet Office: https://www.gov.uk/government/organisations/behavioural-insights-team, Accessed April 30, 2014

[8] See, Center for Strategic Analysis, Ministry of the Republic of France: http://www.strategie.gouv.fr/blog/ , Accessed April 30, 2014

[9] See, Danish Nudging Network: http://www.inudgeyou.com/dnn/, Accessed April 30, 2014

[10] See, "Strengthening Federal Capacity for Behavioral Insights", White House Office of Science and Technology: http://www.foxnews.com/politics/interactive/2013/07/30/behavioral-insights-team-document/, July 30, 2013

[11] See "Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials", UK Cabinet Office Behavioral Insights Team, June 14, 2012

[12] See, "Strengthening Federal Capacity for Behavioral Insights", White House Office of Science and Technology: http://mail.sjdm.org/pipermail/jdm-society/2013-July/005838.html , July 30, 2013

[13] See, Speech by Martin Wheatley, Chief Executive, Financial Conduct Authority (FAC), at City Week 2014 in London " International Financial Services in the Post-Reform World": http://www.fca.org.uk/news/speeches/ethics-and-economics , March 4, 2014

[14] Thaler, Richard H. and Sunstein, Cass R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[15] Iyengar, S.S., Huberman, G., Jiang, W. (2004): "How much choice is too much: determinants of individual contributions in 401K retirement plans." In: Mitchell, O.S., Utkus, S. (Eds.), Pension Design and Structure: New Lessons from Behavioral Finance. Oxford University Press, Oxford, pp. 83–95

[16] Levav, J., Heitmann, M., Herrmann, A., Iyengar, S. (2010): "Order in Product Customization Decisions: Evidence from Field Experiments", Journal of Political Economy, 2010, vol. 118, no. 2

[17] Danziger S, Levav J, Avnaim-Pesso L. (2011): "Extraneous factors in judicial decisions." Proc. Natl. Acad.Sci. USA 108:6889–92

[18] Park CW, Jun SY, MacInnis DJ.(2000): "Choosing what I want versus rejecting what I do not want: an application of decision framing to product option choice decisions". J. Mark. Res. 37:187–202

[19] Johnson EJ, Goldstein D. (2003): "Do defaults save lives?" Science 302:1338–39

[20] Thaler, Richard H. and Sunstein, Cass R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[21] Meredith M, Salant Y. (2011): "On the causes and consequences of ballot order effects." Work. Pap.,Univ. of Pennsylvania

[22] Johnson EJ, Goldstein D. (2003): "Do defaults save lives?" Science 302:1338–39

[23] Madrian BC, Shea DF. (2001): "The power of suggestion: inertia in 401(k) participation and savings behavior." Q. J. Econ. 116:1149–87

[24] Choi JJ, Laibson D, Madrian BC, Metrick A. (2004): "For better or for worse: default effects and 401(k) savings behavior." In Perspectives on the Economics of Aging, ed. DAWise, pp. 81–125. Chicago:Univ. Chicago Press

[25] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[26] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[27] Bertrand M, Karlan D, Mullainathan S, Shafir E, Zinman J. (2010): "What's advertising content worth? Evidence from a consumer credit marketing field experiment" Q. J. Econ. 125:263–306

[28] Thaler RH, Benartzi S. (2004): "Save More Tomorrow™: using behavioral economics to increase employee saving." J. Polit. Econ. 112:S164–87

[29] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[30] Bettinger, E., Long, B., Oreopoulos, P., Sanbonmatsu, L. (2011): "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment", July 2011

[31] See, "Strengthening Federal Capacity for Behavioral Insights", White House Office of Science and Technology: http://mail.sjdm.org/pipermail/jdm-society/2013-July/005838.html , July 30, 2013

[32] See, "Behavior Change and Energy Use", Behavioral Insights Team, UK Cabinet Office, July 2011

[33] Tversky, A., Kahneman, D. (1974): "Judgment under Uncertainty: Heuristics and Biases", Science, New Series, Vol. 185, No. 4157. (Sep. 27, 1974), pp. 1124-1131

[34] Tversky, A., Kahneman, D. (1979): "Prospect Theory: An Analysis of Decision under Risk", Econometrica , 47(2), pp. 263-291, March 1979

[35] Locke, P., Mann, S. (2003): "Professional Trader Discipline and Trade Disposition"

[36] Thaler, R.H. (1980): "Toward a Positive Theory of Consumer Choice", Journal of Economic Behavior and Organization I (1980)39-60.

[37] Thaler, Richard H. and Sunstein, Cass R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[38] See, "Behavior Change and Energy Use", Behavioral Insights Team, UK Cabinet Office, July 2011

[39] Herley, C. (2009): "So Long, and No Thanks for the Externalities: the Rational Rejection of Security Advice by Users" Proceedings of the New Security Paradigms Workshop (NSPW), pp. 133-144

[40] Tversky, A., Kahneman, D. (1981): "The Framing of Decisions and the Psychology of Choice", Science, New Series, Vol. 211, No. 4481. (Jan. 30, 1981), pp. 453-458

[41] Nunes, JC, Boatwright, P. (2004): "Incidental Prices and their Effect on Willingness to Pay," *Journal of Marketing Research*, 41 (4) 457-466.

[42] Thaler, Richard H. and Sunstein, Cass R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[43] Englich, B., & Mussweiler, T. (2001): "Sentencing under uncertainty: Anchoring effects in the courtroom",Journal of Applied Social Psychology,31, 1535–1551.

[44] Mussweiler, T., Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136-164.

[45] Sherman, S.J. (1980): "On the self-erasing nature of errors of prediction." Journal of Personality and Social Psychology, 39.211-22]

[46] Greenwald, A. G., Carnot,C.G., Beach,R.,Young,B.(1987): "Increasing voting behavior by asking people if they expect to vote" Journal of Applied Psychology, 72,315-318.

[47] Berger, J.,Meredith, M., Wheeler, S. (2008): "Contextual priming: Where people vote affects how they vote" PNAS

[48]Leventhal, H., Singer, R., Jones, S. (1965): "Effects of Fear and Specificity of Recommendation Upon Attitudes and Behavior", JPSP, *1965*, 2, 20-29.

[49] Slovic, P., Kunreuther, H., White, G. F. (1974): "Decision processes, rationality and adjustment to natural hazards". In G. F. White (Ed.), *Natural hazards: Local, national, global* (pp. 187-205). New York: Oxford University Press. Reprinted as chapter 1 of P. Slovic (Ed.), *The perception of risk* (pp. 1-31). London: Earthscan Publications, 2000.

[50] Shariff AF, Norenzayan A. (2007): "God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game" Psychological Science. 18:803–9

[51] Randolph-Seng B., Nielsen ME. (2007): "Honesty: one effect of primed religious representations". International Journal of Psychology and Religion 17:303–15

[52] Pichon I, Boccato G, Saroglou V. (2007): "Nonconscious influences of religion on prosociality: a priming Study". Eur. J. Soc. Psychol. 37:1032–45

[53] Payne C (2010): "Personal communication" New Mexico State University College of Business.

[54] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[55] See, Volkswagen's Piano Staircase, www.thefuntheory.com

[56] Peeters, M., Megens, C., den Hoven, E., Hummels, C., Brombacher, A. (2013): "Social Stairs: taking the Piano Staircase towards long-term behavioral change"

[57] See, Policy Options Listen to your Heart, http://www.irpp.org/en/po/nudge-experiments-in-human-nature/listen-to-your-heart/

[58] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[59] Gilovich, T. (1991): "How we know what isn't so: The fallibility of human reason in everyday life". New York: The Free Press. ISBN 0-02-911706-2

[60] See, Gambler's Fallacy: https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Gambler_s_fallacy.html

[61] See, Arizona Daily Wildcat, "Campus Health surveys frats on binge drinking": http://wc.arizona.edu/papers/97/122/01_4.html, March 29, 2004

[62] See, University of Arizona, "A Campus Case Study in Implementing Social Norms and Environmental Management Approaches" http://www.socialnorms.org/pdf/arizonaguidetoalcoholprevention.pdf

[63] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[64] Gino, F., Sharek, Z., Moore, D.A. (2011):"Keeping the Illusion of Control under Control: Ceilings, Floors, and Imperfect Calibration"

[65] Dunn, D. S., Wilson, T. D. (1990): "When the stakes are high: A limit to the illusion-of-control effect". *Social Cognition, 8*(3), 305-323.

[66] Giné, X., Karlan, D., Zinman, J. (2010): "Put your money where your butt is: a commitment contract for smoking cessation". American Economic Journal – Applied Economics 2(4): 213–235.

[67] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[68] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[69] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[70] See, www.stickk.com

[71] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[72] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[73] Cialdini, R.B., Demaine, L.J., Sagarin, B.J., Barrett, D.W., Rhoads, K., Winter, P.L. (2006): "Managing social norms for persuasive impact"

[74] Sherif, M. (1936): "Group Norms and Conformity"

[75]Asch, S. E. (1951): "Effects of group pressure upon the modification and distortion of judgment". In H. Guetzkow(ed.) *Groups, leadership and men*. Pittsburgh, PA: Carnegie Press.

[76] See, "Windshield pitting incidents in Washington reach fever pitch on April 15, 1954": http://www.historylink.org/index.cfm?DisplayPage=output.cfm&File_Id=5136

[77] Strauss, R., Bacon, S.D. (1953): "Drinking in college," New Haven, CT: Yale University Press

[78] Wechsler, H., Davenport, A., Dowdall, G., Moeykens, B., Castillo, S. (1994): "Health and behavioral consequences of binge drinking in college"

[79] Wechsler, H., Nelson, T.F. (2000): "Binge Drinking and the American College Student: What's Five Drinks?"

[80] See, Arizona Daily Wildcat, "Campus Health surveys frats on binge drinking": http://wc.arizona.edu/papers/97/122/01_4.html, March 29, 2004

[81] See, University of Arizona, "A Campus Case Study in Implementing Social Norms and Environmental Management Approaches" http://www.socialnorms.org/pdf/arizonaguidetoalcoholprevention.pdf

[82] Thaler, R.H., and Sunstein, C.R. (2008): "Nudge: Improving Decisions about Health, Wealth, and Happiness", Penguin Books, 2008

[83] Redhead, K. (2008): "A behavioral model of the Dot.com Bubble and Crash", Economics, Finance, and Accounting Applied Research Working Ppaer Series, Coventry University

[84] See, "Behavior Change and Energy Use", Behavioral Insights Team, UK Cabinet Office, July 2011

[85] See, Making Inroad, Policy Options: http://www.irpp.org/en/po/nudge-experiments-in-human-nature/making-inroads/ ;

[86] See, Nudge database, http://economicspsychologypolicy.blogspot.com/2013/03/nudge-database_3441.html

[87] See, "Applying Behavioral Insights to reduce fraud, error, and debt", Behavioral Insights Team, UK Cabinet Office, February 2012

[88] See, Prince George's author of steamy fiction tops list of Maryland's tax cheats, comptroller says http://www.washingtonpost.com/local/md-politics/prince-georges-author-of-steamy-fiction-tops-list-of-marylands-tax-cheats-comptroller-says/2014/01/27/8d10cc06-879b-11e3-a5bd-844629433ba3_story.html, Accessed April 24, 2014

[89] See, Nike Grid, https://www.facebook.com/NikeGrid

[90] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[91] See, "Applying Behavioral Insight to Health", Behavioral Insights Team, UK Cabinet Office, December 2010

[92] See, Beat the Street, http://beatthestreet.me/about

[93] See, "Facebook experiment boosts US voter turnout": http://www.nature.com/news/facebook-experiment-boosts-us-voter-turnout-1.11401 , Nature, September 2012

SRI International

Proposed Research Agenda for Cybereconomic Incentives

March 21, 2014

Prepared for:
Dr. Joseph Kielman
Cyber Security Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Lucien Randazzese, Ph.D., Roland Stephen, Ph.D., and Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development
David Balenson, Ulf Lindqvist, Ph.D., and Zachary Tudor
Computer Science Laboratory

Primary Contact:
Lucien Randazzese, Ph.D.
SRI International
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
lucien.randazzese@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

Introduction

Facing threats from cyber attacks that could disrupt the nation's power, water, communication and other critical national systems, the President issued Executive Order (EO) 13636 on *Improving Critical Infrastructure Cybersecurity*,[1] and on the same day Presidential Policy Directive (PPD) 21 on *Critical Infrastructure Security and Resilience*.[2] These policy documents give the Department of Homeland Security (DHS) an overall coordinating role in pursing the cybersecurity objectives outlined in each document, and directed the National Institute of Standards and Technology (NIST) to work with stakeholders to develop a voluntary framework for reducing cyber risks to critical infrastructure. Part of the work invested in responding to the EO/PPD and developing the NIST framework has been effort by the Departments of Homeland Security, Commerce, and Treasury to identified potential incentives for infrastructure owners and operators to adopt the framework.

The attention to incentives with specific respect to NIST framework adoption reflects growing recognition of the critical role incentives play broadly in cybersecurity. Experience shows that motivated actors – well-intentioned or otherwise – are able to circumvent technical approaches to cybersecurity through manipulation of incentives and human behavior. Accordingly, it follows that cybersecurity can be bolstered by means which address incentive and behavioral considerations. Such approaches to cybersecurity would seek to change the decision-making environment of system developers, vendors, service providers, owners and operators, end-users, and cybercriminals.

This document outlines a proposed cybereconomic incentives (CEI) research agenda to guide the DHS Science and Technology Directorate (S&T) Cyber Security Division (CSD) in identifying and supporting areas of CEI research that will advance CSD's mission of enhancing the security and resilience of the nation's critical information infrastructure.

While the standard microeconomic approach to incentives and behavior provides a useful starting point for analysis – people do often calculate, or attempt to calculate, the gains and losses associated with their choices – in practice there are limits to anybody's ability to reason about the world. Examples abound in real life of this "bounded rationality" by which people seem to choose without perfect (or sometimes any) regard for the costs they will incur or the payoffs they will receive. A holistic approach to researching cybereconomic incentives needs to go beyond the construction of economic cost-benefit models, and leverage knowledge gained from other social science research on incentives and behavior.

---

[1] The White House, *Improving Critical Infrastructure Cybersecurity*, Executive Order 13636, February 12, 2013.

[2] The White House, *Critical Infrastructure Security and Resilience*, Presidential Policy Directive 21, February 12, 2013.

A better understanding of the incentives and behavior of individuals and organizations involved in cybersecurity promises to enhance the government's ability to address two of the principal challenges in securing critical infrastructure:

1. Ensuring that the individuals and organizations that build, deploy, use, and defend these critical assets are incented to make the best decisions with respect to their security.
2. Creating disincentives to attack the nation's infrastructure on the part of malicious entities that might desire to do so.

A large body of research addresses these cybereconomic incentives challenges. As an input to the proposed research areas that follow, the cybereconomic incentives research has been reviewed in detail and summarized in the SRI report, "Literature Review: Current Research in Cybereconomics."[3] The proposed research areas are intended to identify a further body of research that builds on and significantly extends the current research. A second literature review, to be provided in a subsequent paper, will focus on behavioral economics and other social science research primarily in domains outside of cybersecurity, with specific attention to real world applications of research findings to influencing incentives across a range of policy challenges.

The proposed research agenda is organized around six broad research categories

- Category 1: The Economics of Cybersecurity Investment Incentives
- Category 2: Individual Incentives & Behavior
- Category 3: Organizational Incentives & Behavior
- Category 4: Attacker Incentives & Behavior
- Category 5: Cyberinsurance and Cyber Liability
- Category 6: Cybereconomic Incentives Data Collection

There are interrelationships among the categories, with some inevitable overlap in coverage. For example, the study of individual behavior (Category 2) might inform certain questions about how organizations behave (Category 3). Category 6 is cross-cutting, and has implications for research in each of the other five categories.

Within each research category, broad areas of research are described and a number of specific exemplar research questions and topics are identified. A total of ten specific research areas are described across the six proposed categories of research. For some research areas, potential frameworks, databases, or other tools whose development would be useful to policy makers are

[3] SRI International, "Literature Review: Current Research in Cybereconomics," January 3, 2013.

also identified. The last section of this document proposes a few topics for consideration as potential early priorities of the research agenda.

Parts of the proposed research agenda identify issues that are addressable in the short term and have a fairly applied orientation. Other areas pose questions that require more fundamental evaluations of human and organizational behavior, and are therefore generally longer-term in focus. Accordingly, in describing the research areas below, the specific questions associated with each are organized according to their short- versus long-term focus. This distinction is at best a general guide, as there may be shorter and longer term paths of research for many of the individual questions.

As a prelude to preparing this research agenda and the associated CEI literature review, SRI International prepared a concept paper entitled "Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives."[4] This concept paper describes a three-component stakeholder map as a framework for understanding how cybereconomic incentives impact the range of entities involved in deploying, using, and defending critical infrastructure:

- Major players, including network operators, Fortune 500 enterprises, federal agencies, etc.
- Entities further down the value chain, including local service providers and partners, tier two/tier three supply chain enterprises, state and local governments, etc.
- Consumers and employees (often one and the same).

The proposed research areas described below address the cybereconomic incentives affecting each of these three stakeholder groups. The research described for Category 2 (Individual Incentives & Behavior) of course bears directly on consumers and employees, and the research outlined for Category 3 (Organizational Incentives & Behavior) has direct implications for major players and their smaller downstream counterparts. Aside from these one-to-one mappings, however, the proposed research areas outlined for other categories also have relevance for all three stakeholder groups.

The six categories of proposed research also align closely to the fields of research described in the literature review, though the correspondence is not exact. In some cases research that was topically similar but methodologically distinct was separated into distinct sections of the literature review. This proposed research agenda is largely agnostic on issues of research methodology, and accordingly aggregates research streams focusing on the same topic. The table below maps the research agenda categories above with the literature review taxonomy of research. It

[4] SRI International, "Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives," December 18, 2013.

also identifies where the three cybereconomic research projects currently being supported by DHS S&T CSD fall in the research agenda category scheme (in the shaded rows of table).

Some of the proposed research areas also relate to specific Technical Topic Areas (TTAs) that were identified in the 2011 DHS Cybersecurity R&D Broad Agency Announcement (BAA).[5] Where relevant, correspondences between the 2011 TTAs and the proposed research areas are identified. Note that TTA #9 was Cyber Economics, and so it is relevant to all of the proposed areas.

Category 1 **The Economics of Cybersecurity Investment Incentives**	**Category 2** **Individual Incentives & Behavior**	**Category 3** **Organizational Incentives & Behavior**	**Category 4** **Attacker Incentives & Behavior**	**Category 5** **Cyberinsurance and Cyber Liability**
Models of Investment in Cybersecurity (Lit Review Section 3)	Empirical Evaluations of Individual Behavior (Lit Review Section 4.2.1)	Empirical Evaluations of Organizational Behavior (Lit Review Section 4.2.2)	Empirical Evaluations of Attacker Behavior (Lit Review Section: 4.2.3)	Cyberinsurance and Cyber Liability Research (Lit Review Section 5)
	Experimental Evaluations of Behavior (Lit Review Section 4.3)	Experimental Evaluations of Behavior (Lit Review Section 4.3, minimal coverage of organizations)		
	Economic Modeling of Behavior (Lit Review Section 4.1)	Economic Modeling of Behavior (Lit Review Section 4.1)	Economic Modeling of Behavior (Lit Review Section 4.1)	
University of Maryland, *Cyber Economics*			**Carnegie Mellon University,** *Understanding and Disrupting The Economics of Cybercrime*	
Category 6: Cross-Cutting Focus on CEI Data Development				
University of Michigan, *Towards a Global Network Reputation System: A Mechanism Design Approach*				

[5] U.S. Department of Homeland Security. *Cybersecurity Research & Development Broad Agency Announcement*, BAA 11-02, Amendment 00014, June 30, 2011.

Category 1: The Economics of Cybersecurity Investment Incentives

Traditional economics research represents some of the earliest CEI research, with significant effort devoted to theoretical modeling of incentives for various participants in the cybersecurity value chain to protect themselves or to commit cybercrime. This research has shed light on the natural free-rider-based disincentives for adequate investment in cybersecurity, and asserts that the public good nature of cybersecurity leads to underinvestment. Two research areas within Category 1 are outlined below: cybersecurity investment patterns and market impact on cybersecurity investment.

1.1 Cybersecurity Investment Patterns

Most of the existing research on cybersecurity investment is theoretical, and often characterized by highly restrictive modeling assumptions. Accordingly it does not provide much insight into actual cybersecurity investment patterns. The challenges to quantifying enterprise cost for information security have been well articulated. Many of the available estimates on spend are very aggregate and come from market intelligence firms, not from scholars.

Research Area #1: Improved understanding of current patterns of investment in cybersecurity

Representative shorter-term research questions:

a) What are actual levels of investment within the overall system, and how is total spend composed?
b) How does spend vary by cybersecurity activity within the Identify-Protect-Detect-Respond-Recover cycle?
c) What organizational and industry variables, such as industry sector, influence investment?
d) What is the impact of spend at the organizational and system level, and what factors affect how spend translates to security performance and outcomes?
e) How do organizations evaluate the return on investment (ROI) on cybersecurity, and what ROIs are being realized in practice?
f) How do investments in cybersecurity behave from an economic perspective; e.g. are they subject to diminishing marginal returns, economies of scale?

Representative longer-term research questions:

g) How do legislation and public policy impact investment?

Potential frameworks, databases, and other tools to be developed:

h) Frameworks and tools for comprehensive quantification of the costs of cybersecurity and cost of data breaches

Characterizing what owners and operators actually spend on cybersecurity and how they make investment decisions will help policymakers evaluate how under- (or over-) protected the nation's critical infrastructure is, how these investments evolve over time in response to changing threats, and what parts of the overall ecosystem are most at risk.

There is opportunity with this research for interdisciplinary cooperation among economists, business scholars, social scientists, and technologists.

1.2 Market Impacts on Cybersecurity Investment

Research into how market pressures affect firm investments in cybersecurity have focused on relatively few industries (mainly healthcare) and been somewhat inconclusive; some results, for example, suggesting firms in more competitive markets actually do a poorer job protecting customer data.

Research Area #2: The impact of market forces on cybersecurity investment and behavior at firms

Representative shorter-term research questions:

a) How do customers react to security breaches?
b) How do customer reactions to breaches, and the extent to which they are negatively impacted by such breaches, affect firm security investment and behavior?
c) Is it true that most cybersecurity investment by private companies is in Respond-Recover rather than Identify-Protect-Detect, and, if so, what market incentives are driving this behavior?
d) What can be learned from how market forces affect private sector investment in other virtuous goals, e.g. environmental protection, employer health and safety?

Representative longer-term research questions:

e) What is the impact on commerce of more burdensome cybersecurity measures? Are there contexts in which more stringent cybersecurity safeguards actually improve the level of commerce conducted in the marketplace?
f) How do industry concentration and structure impact cybersecurity decisions?
g) How do market forces affect innovation incentives among cybersecurity providers?

Government effort to incent either participation in voluntary frameworks or make legally mandated cybersecurity investment must start with an understanding of how market forces are already affecting these activities and investments. It is in part the inadequacy of market forces to incent optimum investment that provides a rationale for a public policy role in improving cybersecurity.

Category 2: Individual Incentives & Behavior

Cybersecurity is as much a human issue as a technological one, and a significant share of the recent research focuses on the subject of actual individual behavior, both from an empirical and experimental perspective. Work to date in this category suggests a range of additional research areas for further investigation on the third set of cybersecurity stakeholders, consumers and employees, with a particular focus on individuals in their role as employees. Three research areas in Category 2 are outlined below: information and asset stewards, insider threats, and the role of trust in cybersecurity behavior.

2.1 Information and Asset Stewards

Recent empirical and experimental research indicates a number of patterns in individual behavior regarding disclosure of personal data. Individuals are:

- Less likely to act offensively online when their real-world identities cannot be hidden
- More likely to disclose data when they possess a perception of control over disclosure, regardless of whether that control actually impacts use of data by others
- More willing to divulge sensitive personal information when told that others have made sensitive disclosures
- Willing to provide sensitive information, regardless of relevance, if asked for it, and when perceived as part of standard practice
- More willing to tolerate cost and inconvenience if the reason for these are believed to be related to cybersecurity

These findings have clear implications for a range of cybersecurity considerations, but don't address how people behave when they serve as the stewards of data and assets they do not own, especially in their role as employees, of either companies or government organizations, with access to proprietary or classified data or to industrial control systems.

Research Area #3: Behavioral mechanisms affecting individuals when they are responsible for the data and data assets of others

Representative shorter-term research questions:

a) What evaluations can be done of existing efforts to promulgate appropriate cybersecurity behavior (e.g., Stop.Think.Connect.), and how do these inform understanding of how behavior can be changed through programmatic approaches?
b) How do official rules and policies regarding cybersecurity behavior impact task performance and productivity?

Representative longer-term research questions:

c) How do individuals behave differently with respect to cybersecurity when they are responsible for data belonging to others?

d) What behavioral mechanisms that affect personal cybersecurity behavior (e.g. the inability to hide one's identity) affect cybersecurity behavior in a professional context?
e) How do official rules and policies regarding cybersecurity interact with individual behavioral tendencies?
f) How do the informal rules, practices, and policies of an organization, as expressed by its culture, interact with behavioral tendencies that operate at the level of the individual?
g) How does the use of corporate cybersecurity technology meant to restrict people's activities affect individual behavioral tendencies?

Questions c) through g) have been identified as longer-term in resolution because of the methodological challenges expected with understanding how people behave, and misbehave, within a formal work environment. Notwithstanding these methodological challenges, understanding how people behave in the context of a formal organization will be critical to better cybersecurity risk management in the private sector and within government.

The 2011 BAA highlighted Usable Security as TTA #3, where the emphasis was on technology practices and policies that ensure security that is easy for users to use and comply with. Questions e) and g) above address how formal practices and technology like usable security interact with behavior to determine actual cybersecurity outcomes.

2.2 Insider Threats

Existing research into insider threats identifies some expected correlates with information security violations by employees, and also highlights how little many organizations do to protect themselves from such threats. Given how significant the insider threat is to both industry and government, the area merits further research attention.

Research Area #4: Insider threat risks and mitigation

Representative shorter-term research questions:

a) What incentives distinguish true insider threat activity from run-of-the-mill poor cybersecurity hygiene by employees?
b) Why have most insider-threat mitigation strategies focused on procedural mechanisms such as training, password and account management, and activity monitoring?

Representative longer-term research questions:

c) What incentive-based mitigation strategies (legal, technological, operational) could potentially move best-practice beyond procedure-based strategies in the mitigation of insider threats?

Insider Threat is TTA #4 from the 2011 BAA. In outlining the nature of insider threats the BAA indicates that they are so challenging to address because the "**Range of behavioral propensities or triggers which lead to malicious action are difficult to understand** within

the context of insider's knowledge, skills, and experience; environment within which an insider is working; security culture of the organization; and potential recruiting methods of those who would do harm to the organization." While difficult to assess, it is precisely these behavioral propensities, and their associated incentives, that this proposed research area is meant to address directly.

A large body of insider threat case study research exists and is available for use as the partial basis of addressing questions a) and b) of Research Area #4.[6] Understanding the interaction between insider threat mitigation approaches and the behavior factors these approaches are meant to address will likely require research be pushed beyond case studies.

2.3 The Role of Trust on Cybersecurity Behavior

Trust plays a broad and central role in achieving effective cybersecurity. Its importance is embedded in the title of the White House Office of Science & Technology Policy's 2011 strategic plan for cybersecurity R&D: *Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program*.[7] To date most of the research on the role of trust in cybersecurity has focused in three areas:

- Consumer trust of online commerce[8]
- Digital provenance
- Hardware-enabled trust

Within economics and game theory, research has focused on how trust emerges between entities after repeated positive exposure. These technical and social science streams of research suggest a number of paths for further research in the area of trust as it relates to cybereconomic incentives.

Research Area #5: The role of trust in cybersecurity

Representative shorter-term research questions:

a) What lessons are there for cybersecurity from online-banking and other relationships in which individuals entrust sensitive data to organizations?

---

[6] cf. Silowash, George et al. "Common Sense Guide to Mitigating Insider Threats, 4th Edition," Software Engineering Institute, Carnegie Mellon University, December 2012.

[7] White House Office of Science & Technology Policy, "*Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program*," December 6, 2011.

[8] cf. Grabner-Kr.autera, Sonja and Kaluschab, Ewald A. "Empirical research in on-line trust: a review and critical assessment," *International Journal of Human-Computer Studies*, 58 (2003), 783–812.

b) How do organizations evaluate the trustworthiness of individuals with access to information technology and industrial control systems, and where and why do these evaluations fail?
c) How does cybersecurity-related trust correlate with group identification?
d) How is group-identification-determined trust correlated with actual trustworthiness/behavior?

Representative longer-term research questions:

e) How do perceptions of trustworthiness (of other individuals and organizations) affect individual decisions and behavior regarding cybersecurity?
f) How can trust be measured for the various elements that affect the security of critical infrastructure:
    - Individuals
    - Organizations
    - Nation states
    - Data
    - Software
    - Hardware
    - Networks
g) How do technical approaches to establishing trust, such as digital provenance and hardware-enabled trust, affect incentives to invest in cybersecurity in other areas? Might they actually reduce incentives for investment by providing a false sense of security?
h) What are the relative advantages (cost and level of security) of trusting the source of data versus independently verifying the validity of data?
i) Can behavior and practices that lead to trust be encouraged through incentives? Can these incentives be created through public policy intervention?

The 2011 BAA highlighted Hardware-Enabled Trust as TTA #11, where the emphasis focused on ways in which cybersecurity could be realized via technology. The emphasis of the proposed Research Area #5 is on how these and related technology considerations interact with human, behavioral, and incentive aspects in determining the cybersecurity of a system. It seeks to understand why stakeholders trust one another, how this trust is established – via technical or social means – and the role of this trust in both improving cybersecurity and also in security failures.

Each of the three research areas proposed for Category 2, Individual Behavior & Incentives, will benefit from application of findings from behavioral sciences done outside of the cybersecurity domain, and from creativity in methodological approach, including experimentation. The experimental research to date has been focused on how people treat their own data and how they behave as stand-alone “economic” entities. If logistic and ethical constraints can be

overcome, “in-situ” experiments involving actual employees within an organization represent a potentially fruitful research approach.

Category 3: Organizational Incentives & Behavior

Outside of the generic economic theorizing, much of the current research on organizational behavior has been focused on the healthcare industry, as the industry must comply with stringent regulations regarding protecting consumer health information.

Current research into the determinants of organizations’ cybersecurity decisions and performance suggest some preliminary findings, much of it in the healthcare field:

- Expertise, resources, and executive commitment affect cybersecurity performance
- Security compliance is influenced by group affiliations and context
- Decisions to improve cyber-related behavior are influenced by comparisons to how well other organizations perform

Given the enormous diversity of organizations that have cybersecurity concerns, much more research is needed in this area.

Research Area #6: Drivers of organizational cybersecurity behavior

Representative shorter-term research questions:

a) How do the characteristics of organizations, such as mission, leadership characteristics, degree of hierarchy, innovativeness, etc., affect decisions and behavior regarding cybersecurity?
b) Does the impact of these organizational characteristics vary with structural factors such as firm size, legal status, industry sector, financial performance, and the organization of subdivisions within organizations?

Representative longer-term research questions:

c) How does organizational behavior affect trust in an organization, either by its employees, other individuals external to the organization, or other organizations?
d) How do behavioral factors at the individual level aggregate to impact decisions and behaviors at the organizational level?

Part of the challenge in understanding how organizations behave comes from their diversity, a fact acknowledged during the socialization of the preliminary NIST framework, which included consideration of sector-specific implementation of the framework. Good cybersecurity policy is unlikely to be once-size-fits-all, and so a better understanding of how organizations systematically differ in their cybersecurity behavior is key to better public policy.

Category 4: Attacker Incentives & Behavior

Some of the most interesting research in cybereconomics has been focused on cybercriminal and attacker behavior.

Some research has been conducted to evaluate the incentives for cybercrime, and yields some interesting early results:

- Attacks tend to be correlated with one another
- Attacker behavior is affected by significant skewness in target value
- Cybercriminals compete with and cheat one another, exacerbating the negative impact of their activities

This domain of research is relatively new and there remain many opportunities for more study.

Research Area #7: Cybercriminal behavior & incentives

Representative shorter-term research questions:

a) Are there non-financial incentives that will motivate so-called white hat hackers to improve cybersecurity beyond participation in open-source bounties?

Representative longer-term research questions:

b) What incentive differences are there for cybercrime that is motivated by financial gain (economic incentives), vandalism (behavioral incentives), and political/military goals (political incentives), and what implications do these differences have for protecting critical infrastructure?

c) How can the incentives behind hacktivism be harnessed to combat cybercrime rather than merely protest?

Potential frameworks, databases, and other tools to be developed:

d) Database or other characterization of markets (prices, supply, location) for stolen information and for cybercriminal “services”

The 2011 BAA highlighted Modeling of Internet Attacks as TTA #6, focused on malware and botnets, mostly from a technological perspective. The questions posed in this proposed Research Area #7 are not technological but behavioral in focus.

This research would benefit from interdisciplinary research conducted by economists, other social sciences, legal scholars and the law enforcement community.

Category 5: Cyberinsurance and Cyber Liability

5.1 Cyberinsurance

Research on cyberinsurance has tended to focus on understanding why robust markets for cyberinsurance have failed to materialized, concluding that this failure is due to the interdependence of security across insurance holders, the correlation of risk across insurance holders, and information asymmetries. Given the theoretical potential for cyberinsurance to improve cyber-related risk management, more work is needed on ways these obstacles may be overcome.

Research Area #8: Stimulating the emergence of cyberinsurance markets

Representative shorter-term research questions:

a) What do insurance markets and buying behavior for risk areas outside of cybersecurity suggest about how cyberinsurance markets might emerge and operate?
b) What is the cost to firms of having limited cyberinsurance options, or of having to self-insure against cyber-related losses?
c) Can cyberinsurance across a sufficiently diverse population of insurance buyers make reinsurance of cybersecurity risk possible?

Representative longer-term research questions:

d) What role can regulation play in developing a broad market for cyberinsurance?
e) What role can public policy play in reducing information asymmetries believed to be likely in potential cyberinsurance markets?

Potential frameworks, databases, and other tools to be developed:

f) Frameworks for cyberinsurance pricing and actuarial data

Current cybereconomics researchers are somewhat pessimistic about the prospects for a meaningful market for cyberinsurance, and yet significant security failures are precisely the sort of economic events that businesses wish to insure against: low odds risk with high and also highly uncertain costs. Significant social welfare might be achieved if a public policy were able to help overcome the obstacles to such a market emerging.

This research would benefit from cooperation with the insurance industry and legal scholars.

5.2 Cyber Liability

As with cyberinsurance, many scholars assessing the prospects of formal cyber liability mechanisms have concluded that although such mechanisms could in theory help protect information technology users, the costs and logistical challenges of identifying victims and evaluating losses make cyber liability impractical. As in the case of cyberinsurance, the potential

for cyber liability to address the externalities inherent in cybereconomic incentives dictates that more research be done.

Research Area #9: Cyber Liability

Representative shorter-term research questions:

a) What has been the impact of significant data breaches on liability faced by companies that have experience them?

Representative longer-term research questions:

b) What potential legal and policy mechanisms would make cyber liability a more practical tool?
c) How do industry differences (e.g. B2B versus B2C) affect the potential for enforceable liability?

Potential frameworks, databases, and other tools to be developed:

d) Frameworks and tools for cyber-related loss measurement

A functioning cyber liability regime would provide policy makers a tool for addressing the fact that stakeholders rarely bear the full consequences of attacks on their information and information infrastructure.

Research in this area would benefit from the input and perspective of the insurance industry and legal scholars.

Category 6: Cybereconomic Incentives Data Collection

The lack of data remains a critical methodological handicap to cybereconomic research. Better data are needed on:

- Organization cybersecurity policies and activities
- The costs associated with cybersecurity
- Security incidents and their outcomes

The disincentives for sharing data are well known, but the benefits to more widely available information would be considerable for researchers, policymakers, and practitioners. This research area is cross-cutting with the other proposed research areas.

Research Area #10: Cybereconomic Incentives Data Collection

Representative shorter-term research questions:

a) What lessons are there from other policy domains, e.g. the Sarbanes–Oxley Act and financial disclosure?

b) What opportunities are there to use IT assets used in actual cybercrimes as sources of CEI data?[9]

Representative longer-term research questions:

c) What is the role of cybersecurity vendors in providing data?
d) What policy mechanisms could encourage or mandate practical cybersecurity information sharing and discloser?
e) Absent disclosure, what creative mechanisms are there to build CEI-relevant databases for use in policy analysis and actual protection?

Potential frameworks, databases, and other tools to be developed:

f) Comprehensive cross-sectional and time-series databases on cybersecurity investments, breaches and breach outcomes (technological, business, reputational)
g) Database on cybercrimes, including nature, target, location
h) Framework for collection of data given limitations on what can be measured, what will be disclosed

Lack of relevant, timely, accurate, and comprehensive data is a key challenge to policy makers trying to shape effective policy, to owners and operators trying to optimize cybersecurity risk management, and to researchers trying to advance understanding of the cybersecurity and CEI landscape. Even modest progress on some of the questions posed above would help each of these groups with their missions.

Note that two of the cybereconomic research projects currently being supported by DHS S&T CSD address the cybersecurity data collection challenge.[10]

[9] Stone-Gross, Brett, et al., in the paper, "The Underground Economy of Fake Antivirus Software." Tenth Workshop on the Economics of Information Security. George Mason University, Fairfax, VA, USA. 14-15 June 2011, report on the analysis of research done with data recovered from servers used to facilitate cybercrime.

[10] The University of Michigan team lead by Mingyah Liu is investigating ways to encourage data sharing for more accurate reputation assessment; The University of Maryland team lead by Lawrence Gordon is planning to survey firms regarding their cybersecurity investments.

Initial Priorities

This section identifies a small number of specific topics for consideration as potential early priorities of the research agenda. Three criteria were used to identify these potential priorities:

- The relative ease with which short-term progress can be made for the proposed topic
- The amount of insight any quick-to-emerge findings would provide policymakers and other researchers
- The potential of these topics to spur interest and research activity in new areas of cybereconomic incentives

Based on these criteria, the following initial priorities are proposed. These priority topics are selected verbatim from those described above, using the same research area numbering and lettering scheme. Note that all of the priority topics but one fall in the shorter-term question category; the one exception focuses on framework and tool development.

From **Research Area #1** (Improved understanding of current patterns of investment in cybersecurity):

a) What are actual levels of investment within the overall system, and how is total spend composed?
b) How does spend vary by cybersecurity activity within the Identify-Protect-Detect-Respond-Recover cycle?
e) How do organizations evaluate the return on investment (ROI) on cybersecurity, and what ROIs are being realized in practice?
h) Frameworks and tools for comprehensive quantification of the costs of cybersecurity and cost of data breaches

From **Research Area #2** (The impact of market forces on cybersecurity investment and behavior at firms):

a) How do customers react to security breaches?
b) How do customer reactions to breaches, and the extent to which they are negatively impacted by such breaches, affect firm security investment and behavior?

From **Research Area #3** (Behavioral mechanisms affecting individuals when they are responsible for the data and data assets of others):

a) What evaluations can be done of existing efforts to promulgate appropriate cybersecurity behavior (e.g., Stop.Think.Connect.), and how do these inform understanding of how behavior can be changed through programmatic approaches?

From **Research Area #5** (The role of trust in cybersecurity):

b) How do organizations evaluate the trustworthiness of individuals with access to information technology and industrial control systems, and where and why do these evaluations fail?

From **Research Area #7** (Cybercriminal behavior & incentives):

a) Are there non-financial incentives that will motivate so-called white hat hackers to improve cybersecurity beyond participation in open-source bounties?

From **Research Area #10** (Cybereconomic Incentives Data Collection):

a) What lessons are there from other policy domains, e.g. the Sarbanes–Oxley Act and financial disclosure?

SRI International

Proposed Research Experiment for Cybereconomic Incentives

May 30, 2014

Prepared for:
Dr. Joseph Kielman
Cybersecurity Division
HSARPA/DHS S&T
joseph.kielman@dhs.gov

Prepared by:
Bincy Ninan-Moses, Roland Stephen, Ph.D., Lucien Randazzese, Ph.D., and Jeffrey Alexander, Ph.D.
Center for Science, Technology & Economic Development

David Balenson, Ulf Lindqvist, Ph.D., and Zachary Tudor
Computer Science Laboratory

Primary Contact:
Lucien Randazzese, Ph.D.
SRI International
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
lucien.randazzese@sri.com

The views and conclusions contained herein are the authors' and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Department of Homeland Security (DHS) or the US government. The work by SRI International was funded by the DHS Science and Technology Directorate (S&T) under contract no. HSHQDC-10-C-00144.

Introduction

This proposed cybereconomic incentives (CEI) experiment represents the culmination of work performed by SRI International (SRI) for the Department of Homeland Security (DHS) Science and Technology (S&T) Directorate's Cybersecurity Division (CSD). SRI produced a number of related research documents for the DHS aimed at understanding the incentives associated with cybereconomic investment and behavior.

In addition to the experiment described in this document, SRI prepared a concept paper that describes a framework for understanding how cybereconomic incentives impact the range of entities involved in critical infrastructure;[1] assessed the extant research in cybereconomic incentives;[2] reviewed research on behavioral economics and its application to policy domains outside of cybersecurity;[3] and prepared a research agenda to guide CSD in identifying and supporting areas of CEI research that will advance CSD's mission of enhancing the security and resilience of the nation's critical information infrastructure.[4]

Identifying technically effective measures to protect data and critical infrastructure do not guarantee that they will be adopted. Adoption will in large part be determined by incentives, both traditional economic incentives and incentives related to the behavioral tendencies of individuals and organizations. DHS is in the process of mapping a long-term research program focused on cybereconomic incentives. A key first step to launching this program is a proof-of-concept research project that will, ideally, encourage adoption of cybersecurity standards and practices, identify new insights into CEI, and demonstrate the usefulness of the envisioned research program.

A growing share of the existing cybereconomic incentives research focuses on the behavioral aspects of cybersecurity, but in most cases considers personal behavior around privacy, and fails to address how people behave when they serve as the stewards of an organization's sensitive data or infrastructure. The subjects of the existing academic research also tend to be students at the universities where the professors conducting the research reside. The proposed experiment will directly address both of these shortcomings by evaluating how organizations respond to a mix of incentives related to cybersecurity, including behavioral-based incentives.

---

[1] SRI International, "Developing a Proof-of-Principle Exercise for Framing & Investigating Cyber Economic Incentives," December 18, 2013.

[2] SRI International, "Literature Review: Current Research in Cybereconomics," January 3, 2014.

[3] SRI International, "Literature Review: Application of Behavioral Research in Public Policy," April 30, 2014.

[4] SRI International, "Proposed Research Agenda for Cybereconomic Incentives," March 21, 2014.

The above mentioned SRI concept paper on cybereconomic incentives describes a three-component stakeholder map as a framework for understanding how cybereconomic incentives impact the range of entities involved in deploying, using, maintaining, and defending critical infrastructure:

- Major players, including network operators, Fortune 500 enterprises, federal agencies, etc.
- Entities further down the value chain, including local service providers and partners, tier two/tier three supply chain enterprises, state and local governments, etc.
- Consumers and employees (often one and the same).

Major players generally are able to invest more in cybersecurity than the smaller entities downstream in the value chain, making these smaller companies potentially easier targets for attack. Indeed, as cogently demonstrated by the widely-covered data breach at Target Corporation, the security shortcomings of smaller entities can be exploited in attacks on larger organizations. For this reason, and because the number of smaller firms available as potential experiment subjects is large, the proposed experiment focuses on small companies.

Experiment Description

Overview

This experiment is intended to evaluate how small and medium businesses (SMBs) involved with the nation's critical infrastructure respond to incentives to improve their cybersecurity. SMB subjects will either:

- Be offered a no-cost assessment of their potential cybersecurity vulnerabilities, conducted by an appropriate external organization; or
- Be directed to a website where they can take a self-assessment; or
- Be offered a combination of external and self-assessment

Whether participants are offered the external or self-assessment, or the combination of the two, will be determined as the cost and complexity of conducting these two alternatives is evaluated. If that evaluation indicates conducting both types of assessment, the two will be used in conjunction to gauge degree of assessment, with the web-based self-assessment indicating a lower level of participation and the external assessment indicating a high level.

Each assessment offer will include an incentive designed to improve subject interest in going through with the assessment, including incentives inspired by behavioral science. A control group will receive the assessment offer with no incentives. Analysis of response rates will highlight the relative effect of the incentives tested.

The experiment is conceived of as comprising two stages:

1. Pilot stage conducted with a narrow set of SMB respondents to evaluate response rates and degree of insight coming from pilot responses
2. Full rollout with a broader set of SMB respondents, during which treatment effect differences across various firmographic variables, such as firm size and industry, are evaluated

Designing and offering a suitable and valuable cybersecurity assessment will require a review of what's offered currently, either from commercial firms or other organizations such as trade or related associations involved in assisting companies with cybersecurity. In order to fulfill the assessments offered in the experiment, SRI will likely need to partner with an external organization. The review of current cybersecurity assessment offerings will help identify potential partners.

The assessment will be developed within the context of the voluntary framework for reducing cyber risks to critical infrastructure, developed by the National Institute of Standards and Technology (NIST). In other words, it will assess how knowledgeable of and compliant with the NIST framework participants currently are. The goals of the proposed experiment are two-fold: first, to encourage adoption of the NIST Framework; and second, to attract greater interest to research of this kind by demonstrating its usefulness in encouraging adoption and in generating new insights into CEI.

Subject Population & Selection

The subject population for this experiment consists of SMBs involved in some material way with critical infrastructure in the U.S. This includes firms involved in designing, building, servicing, maintaining, or protecting this infrastructure. The exact range of company size in-scope will depend on the total number of companies in the population in question, and ideally will include a relatively large lower bound (i.e. exclude so-called mom-and-pop shops). In identifying subjects, we plan to work with existing industry trade groups and associations to facilitate participation of their members. As needed we will also rely on commercially available marketing lists to fill gaps.

The pilot state of the experiment will be focused on a single industry sector in order to minimize the number of extraneous influences on results. This initial sector will be selected according to the following criteria:

- Sector considered a component of critical national infrastructure
- Includes a large number of member companies that are SMBs
- These SMBs engage in routine and meaningful cyber-interaction with the large companies in the sector

These criteria are meant to ensure that the SMBs included in the experiment interact with key entities in the value chain and are involved in deploying, using, servicing, or defending critical infrastructure. Tentatively, and based on these criteria, the automotive, energy, and financial services sectors are likely candidates for the pilot phase target sector.

Treatment Groups

The SMB subjects of the proposed experiment will be divided into 5 treatment groups and one control group. All subjects will be offered the same cybersecurity assessment. The subjects in the control group will receive the assessment offer without any additional incentive. The subjects in each of the 5 treatment groups will receive the assessment offer plus an additional incentive to agree to the assessment, as described in Table 1 below. These incentives are meant to test the relative impact and effectiveness of behavioral versus economic based incentives. Two of the treatment groups, 4A and 4B, are devoted to measuring the same behavioral impact (loss aversion), one with loss emphasis, the other with gain emphasis.

The proposed incentives offered to the treatment groups listed in Table 1 cover 3 of the 6 concepts in behavioral science around which the behavioral research literature review was organized: loss aversion, anchoring, and peer influence. The experiment does not cover choice fatigue, representativeness, or the illusion of control. It would be impractical to design a single experiment that evaluated too broad a range of factors. The three behavioral factors selected here lend themselves to evaluation in the type of experiment proposed, i.e. one in which an offer is made, subject to some perceived cost and benefit.

Subject Group	Treatment	Treatment Description
Control Group	• *n/a*	• *Simple letter offering cybersecurity assessment, used to calibrate response rates of treatment group*
Treatment Group 1	• Financial incentive	• Assessment offer plus material financial incentive
Treatment Group 2	• Peer influence	• Assessment offer plus positive indication of strong peer participation in assessment
Treatment Group 3	• Anchoring	• Assessment offer plus highlighted reference to recent, high impact breaches (but not impact of those breaches), or some other suitable anchoring mechanism
Treatment Group 4A	• Loss aversion (loss emphasis)	• Assessment offer plus highlighted reference to estimated financial losses at stake in absence of a cybersecurity assessment
Treatment Group 4B	• Loss aversion (gain emphasis)	• Assessment offer plus highlighted reference to estimated financial performance gains to firms that have conducted a cybersecurity assessment

Table 1: Treatment Group Descriptions for Proposed CEI Experiment

Preparation & Staging

As described above, a focused pilot of the experiment will be run in a single sector. The full experiment will be conducted across multiple sectors. The total number and mix of companies included in the full rollout of the experiment will depend on results from the pilot and the estimated cost of fulfilling the assessments and other incentives. Ideally the scope of the second phase will be broad enough to evaluate treatment effect differences across various firmographic variables such as firm size and industry.

In order to design an assessment offer and set of treatments that are appropriately targeted to the subjects under consideration, the experiment team will conduct interviews with selected in-scope SMBs to assess their current needs, attitudes, and understanding of cybersecurity risks and the benefits of an assessment. This information will be used to design the assessment offer invitation and craft the language best suited to convey the various behavioral incentives. In describing the security assessment to participants, language will be used to make the process sound as appealing and non-threatening as possible. For example, loaded terms like audit (as in cybersecurity audit) will be avoided. The interviews will also be used to evaluate the type and size of financial incentive (Treatment Group 1) to be used.