CHALMERS | GÖTEBORG UNIVERSITY

Technical Report No. 08-20

A Prospect Theory approach to Security

VILHELM VERENDEL

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY/
GÖTEBORG UNIVERSITY
Göteborg, Sweden, 2008

Abstract

The correct control of security often depends on decisions under uncertainty. Using quantified information about risk, one may hope to achieve more precise control by making better decisions. We discuss and examine how Prospect Theory, the major descriptive theory of risky decisions, predicts such decisions will go wrong and if such problems may be corrected.

1 Can security decisions go wrong?

Security is both a normative and descriptive problem. We would like to normatively follow how to make correct decisions about security, but also descriptively understand where security decisions may go wrong. According to Schneier [1], security risk is both a subjective feeling and an objective reality, and sometimes those two views are different so that we fail acting correctly. Assuming that people act on perceived rather than actual risks, we will sometimes do things we should avoid, and sometimes fail to act like we should. In security, people may both feel secure when they are not, and feel insecure when they are actually secure [1]. With the recent attempts in security that aim to quantifying security properties, also known as *security metrics*, we are interested in how to achieve correct metrics that can help a decision-maker control security. But would successful quantification be the end of the story? The aim of this paper is to explore the potential difference between *correct* and *actual* security decisions when people are supposed to decide and act based on quantified information about risky options. If there is a gap between correct and actual decisions, how can we begin to model and characterize it? How large is it, and where can someone maybe exploit it? What can be done to fix and close it? As a specific example, this paper considers the impact of using *risk* as security metric for decision-making in security. The motivation to use risk is two-fold. First, risk is a well-established concept that has been applied in numerous ways to understand information security [2, 3, 4, 5, 6] and often assumed as a good metric. Second, we believe that it is currently the only well-developed reasonable candidate that aims to involve two necessary aspects when it comes to the control of operational security: asset value and threat uncertainty. Good information security is often seen as risk management [7], which will depend on methods to assess those risks correctly. However, this work examines potential threats and shortcomings concerning the usability of correctly quantified risk for security decisions.

Our basic conceptual model to understand decision-making for security is as follows, similar to [8]: in this paper, we consider a *system* that a *decision-maker* needs to protect in an *environment* with uncertain threats. Furthermore, we also assume that the decision-maker wants to maximize some kind of *security utility* (the utility of security controls available) when making decisions regarding to different security controls. These different parts of the model vary greatly between different scenarios and little can be done to model detailed security decisions in general. Still, we think that this is an appropriate framework to understand the need of security metrics. One way, maybe often the standard way, to view security as a decision problem is that threats arise in the system and environment, and that the decision-maker needs to take care of those threats with available information, using some appropriate cost-benefit tradeoff. However, this common view overlooks threats with faults that are *made by the decision-maker*. We believe that many security failures should be seen in the light of limits (or potential faults) of the decision-maker when she, with best intentions, attempts to achieve security goals (maximizing security utility) by deciding between different security options.

We loosely think of correct decisions as maximization of utility, in a way to be specified later.

Information security is increasingly seen as not only fulfillment of Confidentiality, Integrity and Availability, but as protecting against a number of threats having by doing correct economic tradeoffs. A growing research into the economics of information security [9, 10] during the last decade aims to understand security problems in terms of economic factors and incentives among agents making decisions about security, typically assumed to aim at maximizing their utility. Such analysis is made by treating economic factors as equally important in explaining security problems as properties inherent in the systems that are to be protected. It is thus natural to view the control of security as a sequence of decisions that have to be made as new information appears about an uncertain threat environment.

Seen in the light of this, and that obtaining security information usually in itself is costy, we think that any usage of security metrics must be related to allowing more *rational decisions* with respect to security. It is in this way we consider security metrics and decisions in the following.

The basic way to understand any decision-making situation is to consider which kind of information the decision-maker will have available to form the basis of judgement. For people, both the available information, but also potentially the way in which it is *framed* (presented), may affect how well decisions will be made to ensure goals. One of the common requirements on security metrics is that they should be able to to guide decisions and actions [11, 12, 13] to reach security goals.

However, it is an open question how to make a security metric usable and ensuring such usage will be correct (with respect to achieving goals) comes with challenges [14]. The idea to use quantified risk as a metric for decisions can be split up into two steps. First, doing objective risk analysis using both assessment of system vulnerabilities and available threats in order to measure security risk. Second, presenting these results in a usable way so that the decision-maker can make correct and rational decisions.

While both of these steps present considerable challenges to using good security metrics, we consider why decisions using quantified security risk as a metric may go wrong in the second step. Lacking information about security properties of a system clearly limits the security decisions, but we fear that introducing metrics do not necessarily improve them, see e.g. [14]. This may be due to 1) that information is incorrect or imprecise, or 2) that usage will be incorrect. This work takes the second view and we argue that even with perfect risk assessment, it may not be obvious that security decisions will always improve. We are thus seeking *properties in risky decision problems* that actually predicts the overall goal - maximizing utility - to be, or not to be, fulfilled. More specifically, we need to find properties in quantifications that may put decision-making at risk of going wrong

In our case, the way to understand where security decisions go wrong is by using how people are predicted to act on *perceived* rather than actual risk. We thus need to use both normative and descriptive models of decision-making under risk. For normative decisions, we use the well-established economic principle of maximizing expected utility. But for the descriptive part, we note that decision faults on risky decisions not only happen in various situations, but has remarkably been shown to happen *systematically* described by models from behavioral economics. In this paper we discuss and examine how the outcome of these models differ and what this difference predicts. The contribution of this paper is summarized as follows

- First, a discussion of rationality and bounded rationality and how these concepts

are important for security decisions, especially when presenting quantitative security risk metrics for people.

- Then, we apply the main descriptive theory of human decisions on risk (Prospect Theory), to see where security decisions are predicted to go wrong using explicit risk as security metric.
- Finally, we investigate the sensitivity of this effect, using numerical studies regarding to correct such problems depending on their sensitivity.

2 Background

Even if one does have a normative model for how risky decisions *should* be made, this says little how such decisions are made in practice. A key challenge in security is to make actual decisions follow the normative standard involving various goals, and it can even be argued that this is a basic reason to do security evaluation.

To study how something may go wrong requires assuming a model of correctness. For risky decisions, we use the standard concept of rationality based on the Expected Utility (EU) principle, initially introduced by Bernoulli [15] and later axiomatized by von Neumann and Morgenstern [16]. The principle is normative in that it prescribes how risky decisions *should* be made for indepdendent decisions, given that we can compute the risk[1] of different options. EU however fails to be descriptive in many ways when it comes to people making decisions in experimental settings.

Deviations from normative risk rules not only happen in various situations, but also to some degree systematically as shown by research in behavioral economics. One of the most prominent models of how peoples risk judgement deviates from EU is Prospect Theory (PT) or its successor Cumulative Prospect Theory, both introduced by Kahneman and Tversky [17, 18], which we will apply to attempt modelling risky security decisions.

We want to specifically model where risk is used as a security metric to study where decisions are predicted to go wrong. In the following, a security decision-maker is faced with a decision by being presented with a number of security *prospects*, where each prospect has a number (one or more) of known-valued outcomes with a probability for each. In the rest of this paper the problem is that one of these prospect has to be chosen over the others. From now on, rationality will be considered for such decisions.

2.1 Rationality

Intuitively a decision picking one from a number of prospects is rational when it gives the best outcome (maximizing utility) for the decision-maker given surrounding constraints. However, most important decisions come not only with a set of fixed outcomes once a decision for an option has been made, but also with uncertainty about outcomes. For such decisions rationality usually means to pick the option which is best *in expecation*. While the expected utility principle has a long history, the modern dominating view of rationality usually relates to Morgernstern and von Neumann who axiomatized the *Expected utility theory*[16]. This showed that if a decision-maker follows a number of simple axioms and has well-known preferences it is shown that there must exist a *utility function* that assigns to each prospect a number to numerically order options

[1] Probabilities of known losses

in such a way that they are ordered in preference. While this lead to a large study and usage of utility functions, this also raised a number of questions to which humans actually act in such a manner. Not surprisingly, this is not always so.

2.2 Bounded Rationality and Prospect Theory

People seem able to quickly make decisions in complex and uncertain environments and often do so quickly without doing complex and deliberate information processing[19][20]. This may be beneficial in with respect to long-term adaption as well as to individual learning with respect to specific environments and is often seen as a combination of both. Ignoring the explanation for such effects, we may also expect to see such simplifying strategies of making decisions to be present in people when it comes to security decisions. These strategies have been largely and systematically studied during the last decades.

The study of how behavior systematically deviates from rationality, in economical and other situations, is the study of *Bounded rationality* that began in the 1950ies [21]. The main finding in bounded rationality has been that human decision-makers often use a set of *heuristics* [20] for their decision-making rather than being fully rational in evaluating what optimal in decisions with regard to the outcomes. These heuristics that can be seen as decision-making shortcuts are believed to rationally reduce the burden on a decision-maker with respect to limited time and resources[2], since they allow more decisions to be made with smaller burden. When said heuristics are used in decisions where they fail they are said to give rise to *bias*. It in such biases that have been largely studied by psychology and economics during the last decades, in the field *Behavioral economics*.

Probably the most well-developed descriptive theory of human decisions using quantified risk is Prospect Theory [17] (PT) (1979) and its successor Cumulative Prospect Theory [18] (1991). PT attempts to describe how people make decisions with quantified risk by modeling decision heuristics directly into the descriptive theory. Three key concepts in PT reflect potential decision bias which differs from normative rational theory. First, decision-makers are *reference-dependent*, meaning that risky prospects are evaluated relative to a reference point rather than to final outcomes. The effect of this subjective viewpoint is known as *framing*, with the reference of the decision-maker affecting how a prospect is qualitatively judged as either a loss or a gain. Second, decisions are *loss-aversive*, meaning that losses are perceived relatively stronger than gains, based on empirical results showing that losses are disproportionally harder to consider when weighted together with gains. Third, probabilities are weighted non-linearly: small probabilities are overweighted while moderate or large probabilities are often underweighted relative to their objective values. The second and third properties attempt to provide explanations understand many non-intuitive effects regarding risk-seeking, risk-aversion and behavior deviating from the purely rational agent. These properties are explicitly modelled using *value* and *weight* (Figures 1,2) functions parametrized to fit empirical results of risky decision-making. A full presentation of PT is outside the context of this paper, we refer the reader to either the Appendix or [22, 23] for good survey and introduction.

[2]Rather than rationality strictly in outcomes

2.3 Risk as a Security Metric

What is commonly known as security metrics still seems to be in a state of ideas about best-practice rather than scientific examination [24] of whether it is rational to use and adopt such metrics. The current state of the field raises the question whether it is really enough with just proposing metrics rather than basing such suggestions on empirical or theoretical validation.

However, the alternative for control of operational security with many decisions under uncertainty is to let experts pick between options using inherently subjective decision criteria [25]. While domain-specific expertise seems the standard way to manage security, this typically does not provide any quantitative methods and measures to understand, control and improve [24] the security risks inherent in different security decisions. One idea behind security metrics is to bridge the gap between domain-specific expert judgement and an application of precise quantitative methods. The goal is to allow precise quantitative evaluation to help guiding the actions of a decision-maker [12], potentially making decisions better.

In general there are many ideas but no strong consensus on what security metrics should be and which properties they need to fulfill their goals. We do not attempt to survey these ideas here. But if security is understood as above, any rational usage of security metrics requires either explicit modelling of gains and losses, or support by empirical work showing the efficiency of letting metrics affect security decisions. This naturally gives two requirements for security in an economic setting: security metrics need to i) provide *precise quantified* indicators of future security performance, and ii) be *rationally usable* with respect to the decision-maker. Now consider two things that may complicate these requirements.

First, deveeloping metrics by measurement of a system in an environment one faces at least two different issues involving uncertainty: i) uncertainty in *measurement*[3] regarding how well one directly or indirectly observes security events that succeed and fail with respect to goals, and ii) uncertainty in an *environment* for how well results can be said to generalize beyond what has been measured in a particular case. With limited information about the future of a system, these uncertainties need to be taken into account. These are major challenges to developing stable metrics for operational situations.

Second, even precise and quantified metrics themselves generally do not come without threats or problems when they are supposed to support decisions (see [14] for a discussion about metrics guiding actions) in a rational way. It has turned out to be a considerable challenge to develop metrics in practice for real-world problems as there are no good established solutions on the horizon. Such metrics are still considered in a stage of lacking both theoretical and empirical evaluation [26] of their efficiency. Our problem in this paper is not how to achieve metrics in the widest sense, but to what extent metrics can be *used* rationally in decision-making. We do not want metrics to provide only perceived confidence but are concerned how they will provide measurable efficiency.

Thus, we see that security metrics needs methods to take uncertainty into account. The only concept that we have found fulfilling these requirements in the literature is to use risk in various ways as a security metric. Formally, the *risk* of an uncertain event means knowing both its probability and the impact of its outcome. Seen in this way, security metrics requires one to model security events and risks in systems involving

[3] For a concrete example: the amount of correct detection by virus/malware detection programs, an IDS, or the confidence one should have in provided expert judgement.

all four parts the basic conceptual model (decision-maker, system, environment and security utility), or to develop security metrics for a decision-maker to perform additional evaluation. We believe that modelling risk in situations with interactions between these is the main challenge to develop good security metrics.

There has actually been no lack of attempts to model risk in complex socio-technical systems, where Probabilistic Risk Assessment [8], decision analysis [27, 28] and dependability [29] are some models that may be used to propose risk metrics. However, little that work has not been directly aimed at security. Some work also ends up involving ways of integrating expert judgement [30, 31], while also relating to potential problems [32, 25] when people are using quantitative methods. One underlying assumption is often that correct modelling will improve systems. Even if such modelling itself is clearly very challenging, in this paper we will *assume* that a decision-maker is provided the result of security risk modelling.

2.4 Related Work

Concepts from Behavioral economics and Prospect theory have been discussed in several places in the security and privacy literature such as [25, 33, 34, 35, 1]. In general, limitations of expert judgement combined with quantitative methods have also been studied in many cases, see [32] for a good introduction on how expert judgement may fail. The work by Schroeder [35] contains experimental work, based on Prospect theory, involving military personel that attempts to repeat several empirical studies made by Kahneman/Tversky by using question-based methods. The author uses questions where the basic structure from previous experimental questions remains - but adapted (on the surface) to a security context. The study claims there is support for bias but that further investigation is needed. Furthermore, some decisions either contain trading off security and operational gains/losses without specifying the measure of security any further, treating security as a goal in itself. Besides not being empirical, two things set the current work apart from [35]. First, this work assumes it is possible to model and estimate costs from security and operational measure into single prospects similar to a monetary sense. Second, we do not yet know of any work that explores bias and reframing systematically around risk that is given as input to security decision-making. This could be used for further hypothesis to investigate Prospect Theory empirically in our setting, complementing intersting initial results from [35].

Among others, the authors in [14] take the view that in order to use metrics well one has to understand how metrics may fail, a view that we precisely examine in this paper for risk as metric.

Using variants of risk as a metric to guide decisions has been proposed in many ways using concepts from economics[2, 3, 5, 6] and Value at Risk-type measures [36] have been proposed to manage security in a financial framework similar to operational risk [4]. Furthermore, risk has been the basis of increasingly technical analysis of how security investments should be done (such as the work started by [37]). Risk metrics span the field between either pure economic risk management and analysis of technical systems, depending on which kind of security system is under consideration. It can be argued that these different methods can all be indirectly used for providing information to risky decisions.

Working with models of perceived risk for non-expert users have been previously discussed, such as in [3]. The authors discuss how risk communication may need to be adapted to the non-expert rather than to experts in certain cases, using experiments with wording, probabilities and various mental models. Further, they state the need to

make mental risk models explicit rather than implicit. Similarly, the issue of assessing system dependability also seems to have ended up examining user confidence [31].

While much work in behavioral economics discusses and reports of the framing effect and human sensitivity [19, 38] to framing with different heuristics, to the best of our knowledge this issue of bounded rationality and framing has not been studied to the degree that it deserves for decision-making and risk assessment in security problems. There seems to be room for applying these tools to understand bad security decisions from new viewpoints, and how judgement may impact security failures.

2.5 Further motivation

Finally, one approach is to simply leave above concerns to decision-makers, where one example is maybe best given by Paté-Cornell in [39], quoted as follows:

In all cases, a probabilistic analysis, of a risk or a decision, should not be performed for people [...] who try to influence them to serve their own purposes, or who have already made up their mind and simply seek justification. [...] If potential users - and the people who are subjected to their decisions - prefer to rely on their insticts, so be it.

Even though such problems are plausible, we take the view that biased usage of information does not have to be left at that. Several arguments can be raised against the view above. First, risk analysis is hardly the only thing that is being used for decision-making in security, even if it is obtained in a correct manner. There may be benefits in actually trying to proactively understand such problems. There may be issues in presenting quantitative information for security decisions that should not be ignored if known beforehand. To avoid acknowledging biased usage of risk analysis may lead to security problems when leading to wrong decisions, like many other usability problems that often turn into security issues. If there is any way to systematically study the phenomena this may also be used to understand the impact of the problem and to suggest possible remedies. When important values are at stake it is not hard to argue for reducing the possibility of wrong decisions.

Furthermore, these problems may obviously be exploited by malicious adversaries who have an incentive to affect the outcome of security decisions. It is important to understand how manipulation of risk perception may happen, which motivates us to study the problem despite that few may be fully unbiased when making risky decisions.

3 Preliminaries

This section presents the modelling of two simple security decision problems. The models consider when a boundedly rational decision-maker is faced with a decision between two prospects a and b regarding to *buy*[4] or *skip* buying protection against security threats.

3.1 Assumptions

Now, the following assumptions are made to get a model suitable for analysis

[4]Here, accepting a prospect containing at least one fixed negative outcome.

- Decision-makers behave as described by Kahneman and Tverskys Cumulative Prospect Theory [18] (denoted as PT). This means that they *make decisions* based on perceived risk and value as described above - so e.g. framing effects may occur.

- Decision-makers have *status quo* as default value reference point, but that may be modified by changing expectations.

- Decision-makers are presented quantified information that is assumed to precisely correspond to the risk in a security problem. We consider where each prospect is presented with negative or positive outcomes and their probabilities.

- The unit for outcomes will be one unit to fit the value function in PT, and *rational* behavior is defined to be linearly dependent on value in expectation (EU). That is, we do not assume normative risk aversiveness, but rather a situation where a decision-maker should normatively be risk-neutral when it comes to risk preferences. This assumption is rather strong, but we feel it may hold when the values at stake are independent and smaller than the base level (status quo of the decision-maker). This is also relevant when one considers repeated but independent decisions (like a large number of different lotteries over time for an entity with relatively large resources).

- Decision-makers are assumed to act solely on the information presented to them with regards to their reference point. We think that this assumption gets more reasonable, combined with the above assumptions, the less the decision-maker has expertise in security issues, such as non-experts with respect to security risks.

3.2 Utility and Prospects

A prospect is a decision option, on shorthand form as follows[18]: a prospect with outcomes $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$ is denoted in shorthand by

$$(x_1, p_1, x_2, p_2, ..., x_n, p_n)$$

If the outcomes $x_1, x_2, ..., x_n$ are exhaustive in that all potential outcome events are listed here, then we require it to be a probability as $\sum_{i=1}^{n} p_i = 1$. Otherwise, by notation, there is an implicit default outcome $(0, p_{n+1})$ with probability $p_{n+1} = 1 - \sum_{i=1}^{n} p_i$.

From now decisions between two prospects are considered. Let a denote the prospect of buying protection to get risk reduction either with certainty or to various degree (examined separately later). Let b denote not buying buying protection only facing certain risk - i.e. accepting a risky outcome instead of either a certain or risky lower absolute loss.

Now, we ask how the normative and descriptive theories differ (no longer prescribe and describe making the same decision) with respect to the actual structure and parameters in decisions. A quick recall of the theories before applying them:
Expected utility: given a prospect $P = (x_1, p_1, x_2, p_2, ..., x_n, p_n)$ where $\sum p_i = 1$, the utility (using the assumptions above) should be[5]

[5] According to the Expected Monetary Value principle [27], which we assume.

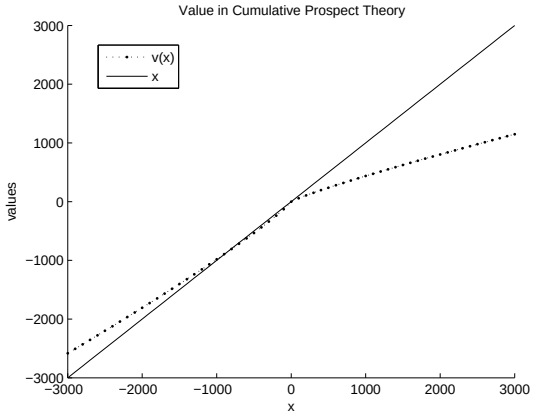

Figure 1: Value in Cumulative Prospect Theory

$$EU(p) = \sum_i p_i x_i$$

Cumulative Prospect Theory: this descriptive theory [18] predicts that preferences between risky prospects are described by a function as

$$V(P) = \sum_{i=1}^{k} \left(w^- \left(\sum_{j=1}^{i} p_j \right) - w^- \left(\sum_{j=1}^{i-1} p_j \right) \right) v(x_i)$$
$$+ \sum_{i=k+1}^{n} \left(w^+ \left(\sum_{j=1}^{i} p_j \right) - w^+ \left(\sum_{j=i+1}^{n} p_j \right) \right) v(x_i)$$

where the value function v and weighting function w are used to evaluate prospects (in terms of positive or negative outcomes, depending on the reference point) as

$$v(x) = \begin{cases} \lambda(x)^\beta & x > 0 \\ -\lambda(-x)^\beta & x < 0 \end{cases}$$
$$w^-(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \text{for negative outcomes}$$
$$w^+(p) = \frac{p^\delta}{(p^\delta + (1-p)^\delta)^{1/\delta}} \text{for positive outcomes}$$

where $\beta, \delta, \gamma, \lambda$ are parameters that have been estimated (from empirical data) to $0.88, 0.69, 0.61, 2.25$ (by regression analysis on a population and picking *median* [18], which is what we use at the moment even though this maybe could be improved). These functions are displayed in Fig 1 and 2. Further brief details can be found in references or in Appendix A.

Initially, we will keep to prospects with negative outcomes. We start with this scenario as work on PT assumes that the status quo is the most natural frame (but we

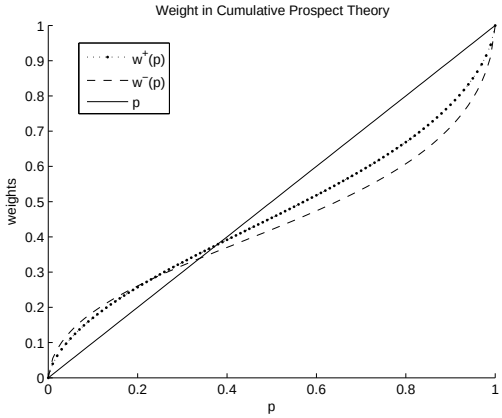

Figure 2: Weighting probabilities in Cumulative Prospect Theory

later examine what the theory predicts when the same prospects are framed differently). That is, initially security decisions are assumed to be made between prospects with all outcomes are perceived as losses (less than or equal to 0), in which case the form of PT for a prospect P simplifies to

$$V(P) = \begin{cases} w^-(p)v(x) \\ \qquad \text{for } P = (x, p, 0, 1-p), x < 0 \\ w^-(p+q)v(x) + w^-(q)(v(y) - v(x)) \\ \qquad \text{for } P = (x, p, y, q, 0, 1-p-q), \\ \qquad y < x < 0 \end{cases}$$

4 Applying Prospect Theory

So given quantified risk analysis - that is, of outcomes and their probabilities, can one find an easy way to decide where such decision-makers are at risk of making wrong decisions? Conversely, where should one want to look for decision failures in order to increase security?

4.1 Failed decisions

Using the previous assumptions decision failures may be stated by constraints as the following:

- **Fail to buy**: we *should* buy protection but Prospect Theory predicts we will not when

$$EU(a) > EU(b)$$
$$V(a) < V(b)$$

- **Fail to skip**: we *should not* buy protection but PT predicts we will when

$$EU(a) < EU(b)$$
$$V(a) > V(b)$$

4.2 Certain protection

In this situation we consider a decision between buying certain protection or facing a fixed loss with a certain probability. To create some intution: finding yourself at risk with the possibility to buy *anti-virus* protection: pay a sum x to get certain protection, or take a risk of facing a much larger loss y with probability p. Formally, a decision-maker has to choose between

- Prospect to buy: $a = (x, 1)$
- Prospect to skip: $b = (y, p)$

We have the two simple prospects $a = (x, 1)$ and $b = (y, r)$ with $y < x < 0$, and want to examine where decisions may differ betwen the best and the actual decision. First, examine where we should buy the protection

$$EU(a) > EU(b)$$
$$\leftrightarrow$$
$$\frac{x}{y} < p$$

We are at risk of not doing so when

$$V(a) < V(b)$$
$$\leftrightarrow$$
$$...$$
$$\leftrightarrow$$
$$\frac{x}{y} > \left(\frac{p^\delta}{(p^\delta + (1-p)^\delta)^{\frac{1}{\delta}}}\right)^{\frac{1}{\beta}}$$

We thus arrive at a relative interval x/y (for the prize and potential loss) where we are at risk of failing to buy:

$$\left(\frac{p^\delta}{(p^\delta + (1-p)^\delta)^{\frac{1}{\delta}}}\right)^{\frac{1}{\beta}} < \frac{x}{y} < p$$

For the reverse conditions of when one should not buy, we are at risk of failing to skip:

$$p < \frac{x}{y} < \left(\frac{p^\delta}{(p^\delta + (1-p)^\delta)^{\frac{1}{\delta}}}\right)^{\frac{1}{\beta}}$$

We will examine these predictions numerically further on in the paper, but first we address where protection is not guaranteed to work - but in itself risky when exposed to a threat.

4.3 Risky protection

Now consider a more realistic situation: protection that is not absolute certain to work but where we (by some method) have a certain confidence in it, information which may be integrated into a risky prospect. This model contains both an uncertainty about whether an attack will happen (a new threat will be realized) and also whether a system will be robust enough to withstand the attack. Threats thus realize themselves in two phases. Using probability, let p be the probability that an attack manifests, and r the conditional probability that an attack succeeds, given protection. The prospects are similar to before, besides that we now have an additional outcome for the prospect of buying protection. Having $y < x < 0$ we have:

- Prospect to buy: $a = (x + y, pr, x, 1 - pr)$
- Prospect to skip: $b = (y, p)$

Lets examine this similar to as before. First, failure to buy. We should choose a when

$$\begin{aligned} EU(a) &> EU(b) \\ &\leftrightarrow \\ pr(x+y) + (1-pr)x &> yp \\ &\leftrightarrow \\ \frac{x}{y} &< p(1-r) \end{aligned}$$

But PT predicts we will not when

$$\begin{aligned} V(a) &< V(b) \\ &\leftrightarrow \\ v(x) + w^-(pr)(v(x+y) - v(x)) &< w^-(p)v(y) \\ &\leftrightarrow \\ &\dots \\ &\leftrightarrow \\ \frac{w^-(p)}{w^-(pr)} &< \left(\frac{x}{y} + 1\right)^\beta + \left(\frac{x}{y}\right)^\beta \left(\frac{1}{w^-(pr)} - 1\right) \end{aligned}$$

The same kind of reasoning can be done for failure to skip. Even if it is possible to find a good closed-form expression for some parameters here ($pr > 0$), instead of diverging into such details, we examine and attempt to illustrate the intervals by numerical study in next section.

4.4 Framing

To examine the sensitivity of framing the decisions in the model are *reframed*, here to two other seemingly natural frames. We aim to to examine what happens with predictions if outcomes in PT are related to different reference points, thus exploiting the

difference in evaluating gains and losses. For new reference points, consider the absolute loss y and the expected loss py. This means keeping the outcomes objectively the same, but assuming the decision-maker perceives some losses as gains, and vice-versa. The idea is to see whether such evaluation predicts reversal of preferences for the predicted failures from the default frame. In the reminder, a reframing of a decision with given parameters is considered *successful* if it is able to predict a corrected decision failure. Evaluation is shown in next section, with the frames look like the following (the sequence of signs in the superscript denote in order whether the outcomes are negative or positive in the new frame)

4.4.1 Certain Protection

We have the potential reframings:

- Case $py < x$:
  - Reference py: $a = (x - py, 1)^{+}$, $b = (y - py, p, -py, 1 - p)^{-+}$
  - Reference y: $a = (-y + x, 1)^{+}$, $b = (0, p, -y, 1 - p)^{0+}$
- Case $py \geq x$:
  - Reference py: $a = (-py + x, 1)^{-}$, $b = (y - py, p, -py, 1 - p)^{-+}$:
  - Reference y: $a = (-y + x, 1)^{+}$, $b = (0, p, -y, 1 - p)^{0+}$:

4.4.2 Risky protection

Next turn to the non-perfect protection mechanisms. As previously seen what complicates things is that buying protection may now have two outcomes, that of successful protection and that of an even larger loss. For the risky protection we have $y + x < y < \{py < x, x < py\}s < 0$. We have the potential reframings:

- Case $py < x$:
  - Reference py: $a = (x+y-py, pr, x-py, 1-pr)^{-+}$,$b = (y-py, p, -py, 1-p)^{-+}$
  - Reference y: $a = (x, pr, x - y, 1 - pr)^{-+}$, $b = (-y, 1 - p)^{+}$
- Case $py \geq x$:
  - Reference py: $a = (x + y - py, pr, x - py, 1 - pr)^{--}$[6], $b = (y - py, p, -py, 1 - p)^{-+}$
  - Reference y: $a = (x, pr, x - y, 1 - pr)^{-+}$, $b = (-y, 1 - p)^{+}$

[6]NB! V reduces to $v(x - py) + w^{-}(pr)(v(x + y - py) - v(x - py))$

5 Numerical Evaluation

This section describes how to use numerical methods to study the impact of the above issues. Two things are studied:

First, to examine how PT predicts that decisions will go wrong. The following method is used here: fix $y = -5000$ (somewhat arbitrarily to study the problem, but it fits the monetary scale to which Prospect Theory was fit [18]) and iterate (discrete) $x \in [y, 0]$ to see what PT predicts for the scale used in PT. For each of the parameters (x,p and also r for risky protection) PT is applied and the predicted outcome is compared with the normative decision prescribed by EU. This is used to see which kinds, with respect to prarameters, of decision problems are predicted to be sensitive to decision failure.

Second, how robust the found decision failures are to the frame of the prospects. This is done by reframing the problems to be expressed from different reference points. *For each* seen buy or skip failure PT is used to evaluate the same parameters when using a different reference-point to see if reframing predicts a reversal in preferences ("correcting" the failure). The new reference points are based on applying PT for the *absolute loss* y and *expected loss* py as above, which seem to be the most plausible frames alternatives.

5.1 Results

The numerical results show that PT predicts decisions to go wrong for some x/y intervals for certain parameters in both problems.

For certain protection, fig 3 shows the boundaries of the x/y intervals where PT predicts failure in decisions. The interval (between the curves) for small probabilities predicts failure to skip (risk-aversiveness), and for larger probabilities for which x/y prediction of failure to buy (risk-seeking).

For risky protection, see fig 4 and 5 for where failures are predicted at all for *some* x/y. For lower and upper boundaries of failure to buy see fig 6 7, and for failure to skip see fig 8 and 9.

For reframing certain protection, see fig 10 and 11. Reframing cases of failure to buy is predicted successful in reversing preferences in 90.8% and 100% for the absolute and expected frames, respectively. Reframing when failure to skip had 0%(!) success rate.

For reframing risky protection, see fig 12-15. For the z-axis, -1 denotes that no failures have been predicted (avilable for successful reframing) in the initial test. The other data points denote how large fraction of decision failures are successfully corrected under another frame. Success rate for failure to skip: absolute frame 13.5%, expected frame 59.2%. Success rate for failure to buy: absolute frame 74.5%, expected frame 74.5%. There is full overlap by the largest amount of success in both cases. These experiments show that there are clearly scenarios where a large fraction of the predicted decision problems can not be fixed by reframing the expressed risk. Knowing the correct numbers may be used to predict deviation from rational security policy.

6 Conclusion

We have considered when quantified risk is being used by people making security decisions. An exploration of the parameter space in two simple problems showed that results from behavioral economics may have impact on the usability of quantitative risk methods. The results visualized do not lend themselves to easy and intuitive explanations, but we view our results as a first systematic step towards understanding security problems with quantitative information.

There have been many proposals to quantify risk for information security, mostly in order to allow better security decisions. But a blind belief in quantification itself seems unwise, even if it is made correctly. Behavioral economics shows systematic deviations of weighting when people act on explicit risk. This is likely to threaten security and its goals as security is increasingly seen as the management of economical trade-offs. We think that these findings can be used partially to predict or understand wrong security decisions depending on risk information. Furthermore, this motivates the study how strategic agents may manipulate, or attack, the perception of a risky decision.

Even though any descriptive model of human decision-making is approximate at best, we still believe this work gives a well-articulated argument regarding threats with using explicit risk as security metric. Our approach may also be understood in terms of standard system specification and threat models: economic rationality in this case is the specification, and the threat depends on bias for risk information. We also studied a way of correcting the problem with reframing for two simple security decision scenarios, but only got partial predictive support for fixing problems this way. Furthermore, we have not found such numerical examinations in behavioral economics to date.

Further work on this topic needs to empirically confirm or reject these predictions and study to which degree they occur (even though previous work clearly makes the hypothesis clearly plausible at least to some degree) in a security context. Furthermore, we think that similar issues may also arise with several forms of quantified information for security decisions.

These questions may also be extended to consider several self-interested parties, e.g. in game-theoretical situations. Another topic is using different utility functions, and where it may be normative to be economically risk-aversive rather than risk-neutral.

With respect to the problems outlined, rational decision-making is a natural way to understand and motivate the control of security and requirements on security metrics. But when selecting the format of information, a problem is also partially about usability. Usability faults often turn into security problems, which is also likely for quantified risk. In the end the challenge is to provide users with usable security information, and even more broadly investigate what kind of support is required for decisions. This is clearly a topic for further research since introducing quantified risk is not without problems. Using knowledge from economics and psychology seems necessary to understand the correct control of security.

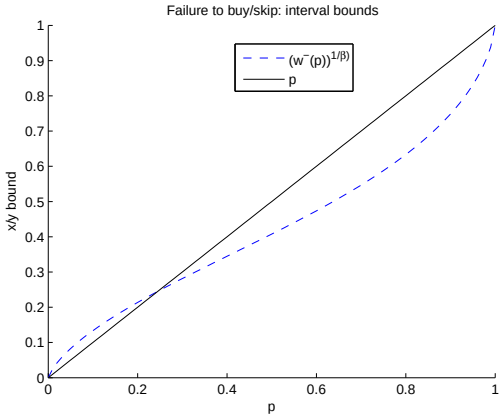

Figure 3: Failures with certain protection

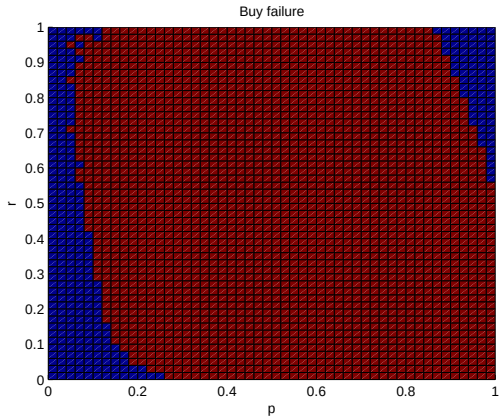

Figure 4: Failures (red) with risky protection

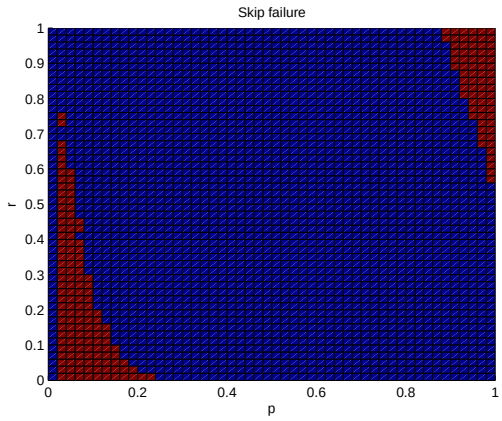

Figure 5: Failures (red) with risky protection

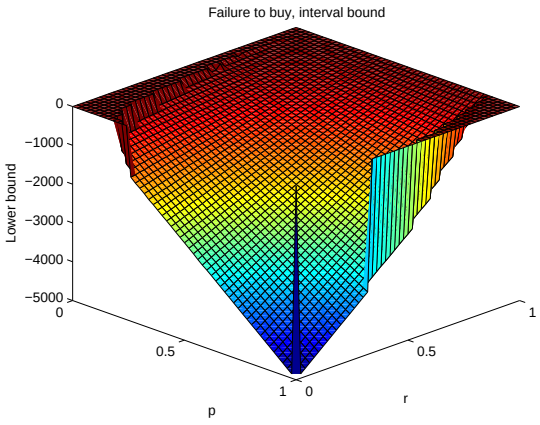

Figure 6: Failure to buy risky protection: lower bound

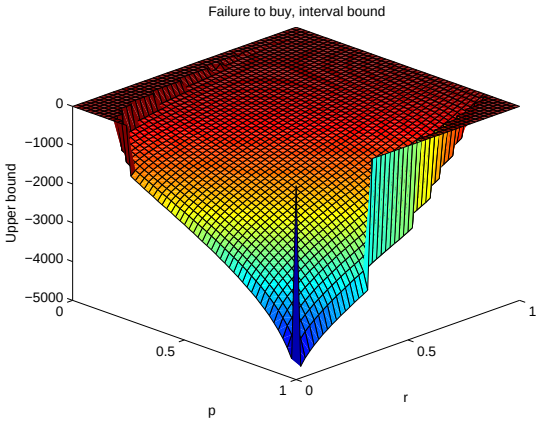

Figure 7: Failure to buy risky protection: upper bound

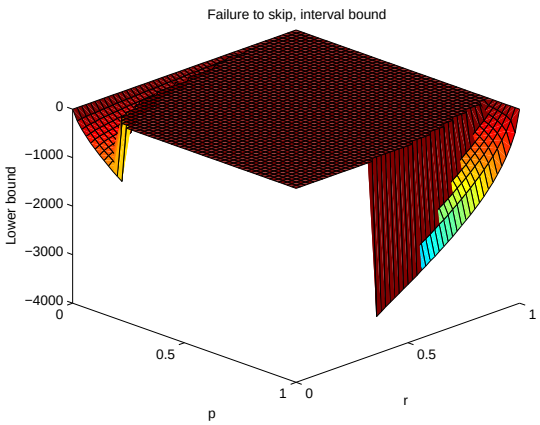

Figure 8: Failure to skip risky protection: lower bound

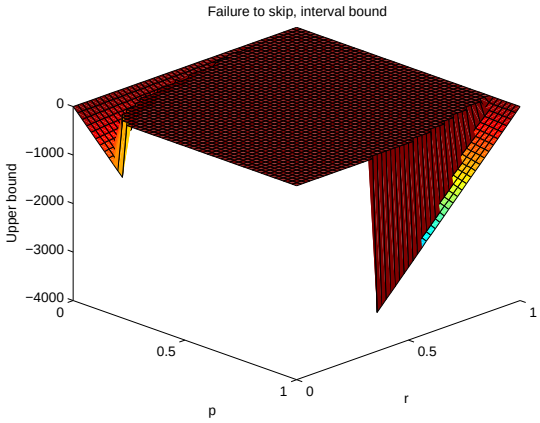

Figure 9: Failure to skip risky protection: upper bound

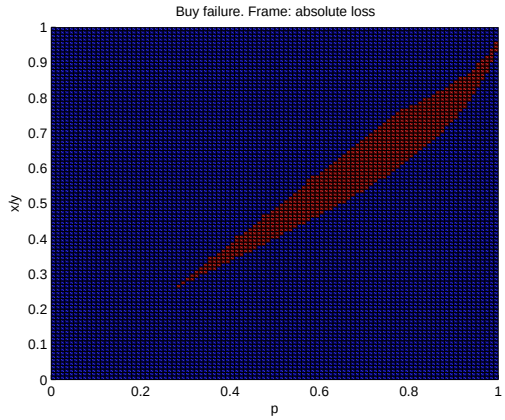

Figure 10: Reframing with certain protection

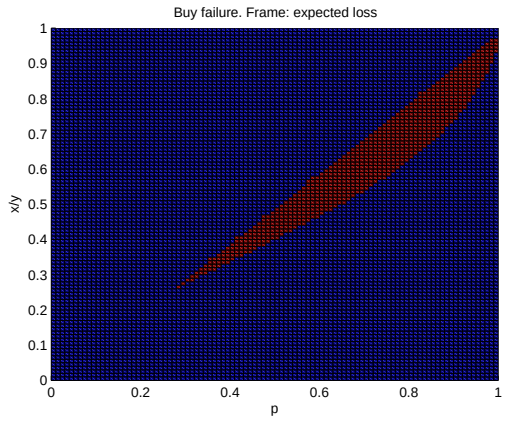

Figure 11: Reframing with certain protection

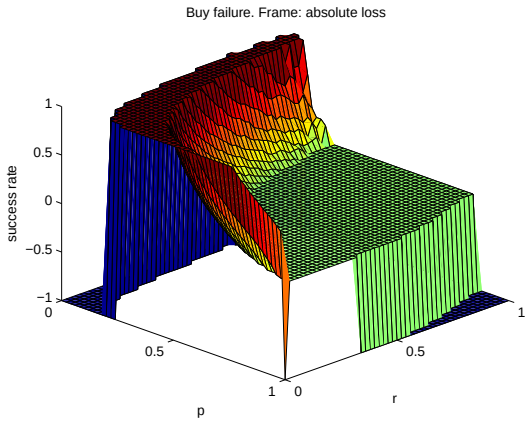

Figure 12: Reframing risky protection (fail to buy)

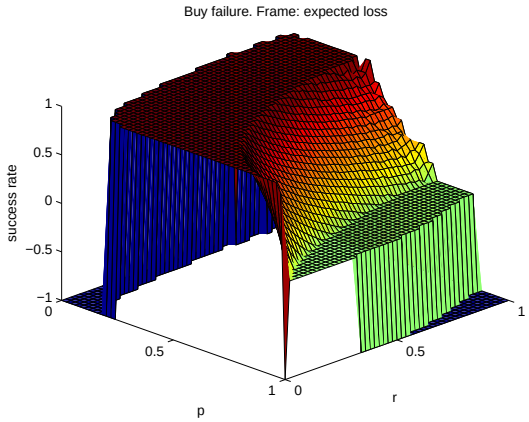

Figure 13: Reframing risky protection (fail to buy)

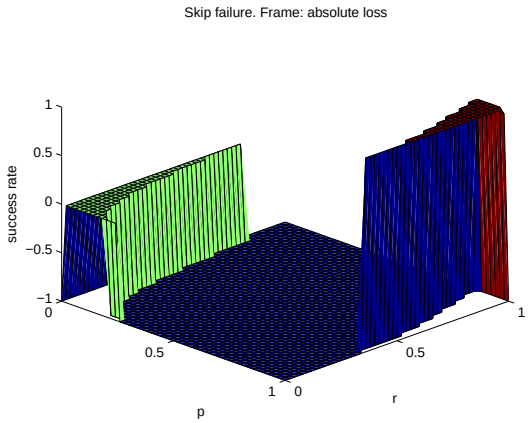

Figure 14: Reframing risky protection (fail to skip)

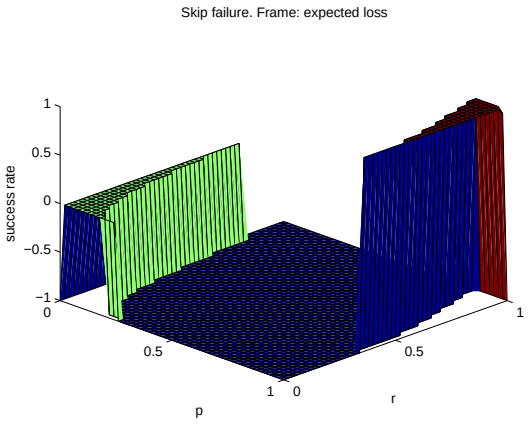

Figure 15: Reframing risky protection (fail to skip)

References

[1] B. Schneier, "The psychology of security," 2007.

[2] S. E. Schechter, "Toward econometric models of the security risk from remote attacks," *Security & Privacy, IEEE*, vol. 3, no. 1, pp. 40–44, 2005.

[3] F. Asgharpour, D. Liu, and J. L. Camp, "Mental models of computer security risks," in *Workshop on the Economics of Information Security*, 2007.

[4] R. Hulthén, "Communicating the economic value of security investments; value at security risk," in *Workshop on the Economics of Information Security*, 2008.

[5] Y. Asnar, R. Moretti, M. Sebastianis, and N. Zannone, "Risk as dependability metrics for the evaluation of business solutions: A model-driven approach," in *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*, 2008, pp. 1240–1247.

[6] "The financial impact of cyber risk," American National Standards Institute (ANSI) / Internet Security Alliance (ISA), 2008.

[7] G. Stoneburner, A. Goguen, and A. Feringa, "Risk management guide for information technology systems," 2002.

[8] V. M. Bier and L. A. Cox, "Probabilistic risk analysis for engineered systems," in *Advances in Decision Analysis - From Foundations to Applications*, W. Edwards, R. F. Miles, and D. von Winterfeldt, Eds. Cambridge, 2007, pp. 279–301.

[9] R. Anderson, "Why information security is hard - an economic perspective," in *Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual*, 2001, pp. 358–365.

[10] R. Anderson and T. Moore, "The economics of information security: A survey and open questions," 2006.

[11] M. Swanson, N. Bartol, J. Sabato, J. Hash, and L. Graffo, "Security metrics guide for information technology systems," NIST, Tech. Rep., 2003.

[12] S. C. Payne, "A guide to security metrics," 2006.

[13] E. Chew, M. Swanson, K. Stine, N. Bartol, A. Brown, and W. Robinson, "Nist performance measurement guide for information security (draft)," NIST, Tech. Rep., September 2007.

[14] J. Hauser and G. Katz, "Metrics: you are what you measure!" *European Management Journal*, vol. 16, no. 5, pp. 517–528, October 1998.

[15] D. Bernoulli, "Exposition of a new theory on the measurement of risk," *Econometrica*, vol. 22, no. 1, pp. 23–36, 1954.

[16] O. Morgenstern and J. Von Neumann, *Theory of Games and Economic Behavior*. Princeton University Press, May 1944.

[17] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–292, 1979.

[18] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, October 1992.

[19] R. Hastie and R. M. Dawes, *Rational Choice in an Uncertain World: The Psychology of Judgement and Decision Making*, 2nd ed. Sage Publications, 2001.

[20] D. Kahneman, P. Slovic, and A. Tversky, *Judgment under Uncertainty : Heuristics and Biases*. Cambridge University Press, April 1982.

[21] H. A. Simon, "A behavioral model of rational choice," *The Quarterly Journal of Economics*, vol. 69, no. 1, pp. 99–118, 1955.

[22] D. Kahneman, "Maps of bounded rationality: Psychology for behavioral economics," *The American Economic Review*, vol. 93, no. 5, pp. 1449–1475, 2003.

[23] D. Kahneman and A. Tversky, *Choices, Values, and Frames*. Cambridge University Press, September 2000.

[24] C. Villarrubia, E. F. Medina, and M. Piattini, "Towards a classification of security metrics," in *WOSIS*, 2004, pp. 342–350.

[25] L. Strigini, "Engineering judgement in reliability and safety and its limits: what can we learn from research in psychology," CSR, Tech. Rep., 1996.

[26] C. Villarrubia and E. F. Medina, "Analysis of iso/iec 17799: 2000 to be used in security metrics." in *Security and Management*. CSREA Press, 2004, pp. 109–117.

[27] H. Raiffa, *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Addison-Wesley, 1968.

[28] W. Edwards, R. F. Miles, and D. von Winterfeldt, *Advances in Decision Analysis: From Foundations to Applications*. Cambridge, 2007.

[29] T. A. Delong, D. T. Smith, and B. W. Johnson, "Dependability metrics to assess safety-critical systems," *Reliability, IEEE Transactions on*, vol. 54, no. 3, pp. 498–505, 2005.

[30] S. Hora, "Eliciting probabilities from experts," in *Advances in Decision Analysis*, W. Edwards, Miles, and D. von Winterfeldt, Eds. Cambridge University Press, 2007.

[31] R. E. Bloomfield, B. Littlewood, and D. Wright, "Confidence: Its role in dependability cases for risk assessment," in *DSN '07: Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 338–346.

[32] E. Yudkowsky, "Cognitive biases potentially affecting judgment of global risks," in *Global Catastrophic Risks*, N. Bostrom and M. Cirkovic, Eds., 2006.

[33] J. J. Gonzales and A. Sawicka, "The role of learning and risk perception in compliance," in *21st System Dynamics Conference*, 2003.

[34] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *Security & Privacy, IEEE*, 2005.

[35] N. J. Schroeder, "Using prospect theory to investigate decision-making bias within an information security context," Master's thesis, Department of the Air Force, 2005.

[36] P. Jorion, *Value at Risk, 3rd Ed.* McGraw-Hill, October 2006.

[37] L. A. Gordon and M. P. Loeb, "The economics of information security investment," *ACM Trans. Inf. Syst. Secur.*, vol. 5, no. 4, pp. 438–457, November 2002.

[38] M. H. Bazerman, *Judgement in Managerial Decision Making*, 5th ed. Wiley, 2006.

[39] E. Paté-Cornell, "Probabilistic risk analysis versus decision analysis: Similarities, differences and illustrations," in *Uncertainty and Risk: Mental, Formal, Experimental representation*. Springer, 2007, pp. 223–242.

[40] D. Prelec, *Compound Invariant Weighting Functions in Prospect Theory*. Cambridge University Press, 2000.

A Values and weights in (Cumulative) Prospect Theory

This section contains a few technical details about prospect theory and the differences between the original [17] and the cumulative version [18]. First, recall the expected value of a prospect P is expressed as

$$EU(P) = \sum_{i=1}^{n} p_i x_i$$

The original version of PT offers an almost similarly elegant and simple formula, for prospects with two outcomes the theory claims there is a function V^* that describes ordering of preferences (as a utility function) as

$$V^*(P) = \pi(p_1)v^*(x_1) + \pi(p_2)v^*(x_2)$$

where the value and weighting functions v^* and π are similar to the functions later used in the cumulative version (Fig 1, 2). But the cumulative version for two prospects (applied in this work) does not appear that simple, at least on the surface. A good question is: Why? First, original prospect theory is limited to describe prospects with only two non-zero outcomes. Second, probabilities are judged slightly different for losses than gains - the original version uses the same weighting function π both for positive and negative outcomes. Third, a somewhat more technical need is to tackle a problem known as stochastic dominance (for more on this, see e.g. [40]) that arises to the non-linearity in probability weighting. Several so-called *rank-based* theories have been proposed to also provide also this third property, where Kahneman/Tverskys version is one of the proposed theories.

Now, Cumulative Prospect Theory proposes the following: given a prospect $P = (x_1, p_1, x_2, p_2, ..., x_n, p_n)$ where $\sum p_i = 1$ and the outcomes $x_1, x_2, ..., x_n$ are increasingly ordered as $x_1 \leq x_2 \leq ... \leq x_k \leq 0 \leq x_{k+1} \leq ... \leq x_n$, then there exists a function v that describes ordering of preferences with a function V (similar to above) as

$$V(P) = \sum_{i=1}^{k} \left(w^- \left(\sum_{j=1}^{i} p_j \right) - w^- \left(\sum_{j=1}^{i-1} p_j \right) \right) v(x_i)$$
$$+ \sum_{i=k+1}^{n} \left(w^+ \left(\sum_{j=1}^{i} p_j \right) - w^+ \left(\sum_{j=i+1}^{n} p_j \right) \right) v(x_i)$$

To obtain the simple form for application in section 3.2 is a matter of applying this formula to specific forms of prospects (e.g. for when all outcomes are negative and so forth).

The value and weighting functions used in this theory are shown in Fig 1, 2 - as introduced in section 3.2. The value function models both the widely accepted principle of diminishing marginal utility (gains and losses), as well as *loss aversiveness* - that gains and losses by equal magnitude do not cancel out. The weighting function describes overweighting for smaller and underweighting of moderate or relatively large (cumulative) probabilities in the formula for PT above.