

Optimal Private and Public Warning Decisions in a Dynamic Context:

An Extension to Pinker's Normative Framework of Security

Navid Ghaffarzadegan

PhD student in Decision and Policy Sciences

Rockefeller College of Public Affairs and Policy, University at Albany, SUNY

David F. Andersen

Distinguished Service Professor

Rockefeller College of Public Affairs and Policy, University at Albany, SUNY

Abstract: A normative approach to security problems can give important insights into what the most proper security policies are for a given set of circumstances. Pinker (2007) proposes a framework for optimizing the short term policies of guard allocation and private and public warning issuance. In this paper we focus on the issuance of public and private warnings by extending Pinker's model to a dynamic setting. We propose a simulation model based on Signal Detection Theory that enables the determination of optimal private and public warning decisions. Through the extended model we show that optimal solutions are sensitive to two major assumptions: sensitivity to false alarms and terrorists' perceptions of public sensitivity. The results show that an underestimation of these effects can result in biases in optimal solutions.

Keyword: security, public warning, private warning, normative decision making

1. Introduction

Issuing private and public warnings are two of the most common defensive policies for any government. Warnings can improve public awareness as well as guards' readiness to face an attack and prevent the consequences. Further, public warning issuance can have a deterrence effect and stop terrorists from implementing a planned attack. On the other hand, extensive issuance of warning can result in social stress. From a normative decision making stand point, this trade off leads to finding a policy for public and private warning issuance that minimizes the total costs of attack and warning issuance.

However, an optimal policy of warning can exist even if we assume no direct cost of warning issuance. Issuance of warning when no attack happens - false alarms - may have significant effects on the social system. False alarms can accumulate in individuals' memory in long run and may result in a loss in public and private sensitivity to warnings - also known as the crying wolf effect. The effectiveness of a warning is in fact not only a function of the accuracy of the warning, but also people's responsiveness (Pate-Cornell 1986) and a future decline in the latter effect can be considered as a cost component. Furthermore, the loss of public sensitivity can feed back to terrorists' behavior and weaken the deterrence effect of warning issuance, resulting in attacks that might not have occurred if terrorists perceived that the public was responsive to warnings. Therefore, the accumulation of false alarms and the information feedback to terrorists about a potentially desensitized public may add to the long term costs of a warning policy and result in a shift in optimal policies.

In his Management Science article, Pinker (2007) studies the effect of short-term responses to security and proposes a framework to estimate the optimal level of guard

allocation and private and public issuance of warnings. While pinker proposes problem definitions and associated equations that lead to a minimization of total cost in each time period, an extension of his work to optimize the long-term cost can be interesting and insightful. A dynamic extension of a policy helps us to examine the long-term side effects of a policy and help policymakers to avoid better-before-worse patterns of behavior (Forrester 1971). Studies on security policies show that this domain of policymaking is prone to decisions that may seem proper in short term but backfire in long term (Sagan 2004, Ghaffarzadegan 2008). This fact increases the need for the examination of security policies in a dynamic context.

In this paper we focus on the issuance of public and private warnings and extend Pinker's model to a dynamic setting. We use Signal Detection Theory (Macmillan and Creelman 1991, Green and Swets 1966; Swets 1991) to model warning decision making in Pinker's model in order to facilitate finding numerical solutions and optimal private and public warning decisions that minimize damage from possible attacks. The dynamic extension helps us to get an in-depth understanding of the effect of false alarms on private and public sensitivity. Through the extended model we show that optimal solutions are sensitive to two major assumptions: public sensitivity to false alarms and terrorists' perception of public sensitivity, whereby an underestimation of these effects results in biases in optimal solutions. In the next sections of the paper we introduce a warning issuance model (section 2), conduct simulation analysis, and examine the results (section 3).

2. Modeling

2.1. Security judgment and warning decision making

We use signal detection theory (Macmillan and Creelman 1991, Green and Swets 1966; Swets 1991) to model judgment and decision making on warning issuance. This framework divides events into two categories, the ones we want to detect (positive events) and others (negative events). In terms of security, we are interested in differentiating the times that we will be attacked (positive events) from the times that we will not be attacked (negative events). A perfect detection leads to proper warning decisions.

In order to detect positive events, from signal detection perspective, a decision maker compares continuous judgment about how suspicious an event looks like (for example, it can be based on reports from the intelligence system) with some thresholds for warning issuance and then makes a decision about whether or not private and/or public warnings should be issued.

As there is always uncertainty in the environment, it is not completely possible to detect positive events, so the distributions overlap. Figure 1 assumes a normal distribution for positive and negative events (over judgment, not time) and shows the distributions overlap. So, wherever we position our decision threshold for public and private warnings there is always a chance that we make a wrong decision.

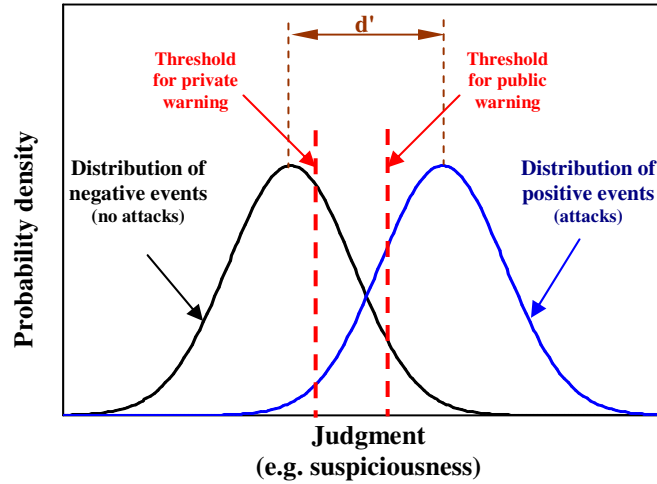


Fig.1: Distribution of positive events (attacks) and negative events (no attack) in signal detection theory

In Figure 1, the distance between the means of the positive and negative event distributions is labeled d' and represents judgmental accuracy which may improve by improving the intelligence systems. The vertical lines are decision thresholds for private and public warnings. In this framework, the proportion of positive events to total events is called the base rate. For a particular case when judgment falls below both of the thresholds then no warning action will be taken. If it falls between the thresholds only private warnings will be issued, and if judgment falls above both of the thresholds, a public warning is issued which is equivalent to issuing both public and private warnings. Mathematically we can say:

$$W_i = \begin{cases} 1 & \text{if } x > C_i \\ 0 & \text{if } x \leq C_i \end{cases} \quad \text{equation 1}$$

Whereby i represents one of the two agents, i.e., $i \in \{private, public\}$. W_i is a warning decision and C_i is the threshold for the i kind of warning. Further, x is judgment about

how severe an event looks like (continues variable). So, for example if $x > C_{public}$ then $W_{public} = 1$, i.e., public warning will be issued.

In any decision making situation, there are four possible outcomes from each agents' perspective. A warning can be issued and an attack may happen or not (true positive and false positive), or a warning may not be issued and again an attack may happen or not (false negative and true negative). We name these outcomes as TP_i , FP_i , FN_i , and TN_i respectively. Errors include positive events that we miss (false negative, i.e., misses), and negative events that we alarm (false positives, i.e. false alarms). As it appears from the figure, without moving the distribution further apart, we cannot decrease both of the errors at the same time. For example, moving the thresholds towards the right increases misses and decreases false alarms, and vice versa. Usually different values/costs are socially considered for each outcome, and therefore, optimal decision thresholds can be found.

As guards are assumed to be aware of public warnings, total of six possible outcomes for private and public warning decision making emerge. Table 1 shows these possible decision outcomes. For simplicity we name them as $O_1 - O_6$. As stated in the Table, terrorists may cancel a plan for attack when public warning is issued. Let's define $R_{terrorists}$ as terrorists' reaction to warning issuance. $R_{terrorists}$ is 1 if a public warning is issued and terrorists respond to it by canceling their attack and otherwise is zero. Interestingly, when it happens, the event maybe perceived by the public as a false alarm. So, we can say $FN_{private} = O_1$, $TP_{private} = O_2 + O_3 \cdot (1 - R_{terrorists})$, $TN_{private} = O_4$, and $FP_{private} = O_5 + O_6 + O_3 \cdot R_{terrorists}$. With a similar logic, for the public sector we have: $FN_{public} = O_1 + O_2$, $TP_{public} = O_3 \cdot (1 - R_{terrorists})$, $TN_{public} = O_4 + O_5$, and $FP_{public} = O_6 + O_3 \cdot R_{terrorists}$.

		warning decision		
		no action	private warning	public warning
state of the world	Plan to attack	O ₁ : - False negative decision (Misses). - Very costly for a society.	O ₂ : - Guards are ready – Public is not ready. - improve in defense effectiveness	O ₃ : - Everyone is ready - improve in defense effectiveness. - Deterrence can occur and attack may be canceled
	No plan to attack	O ₄ : - True negative decision	O ₅ : - Over-reaction in guards. - May result in loss in sensitivity to warning in guards	O ₆ : - Over-reaction in public. - May result in stress. - May result in a loss in sensitivity to warnings

Table 1: Six possible outcomes for warning decisions

Let's assume that terrorists cancel their plan to attack (positive event) when a public warning is issued with the probability of $p_{canceling}$. Then damage from an attack can be formulated as:

$$\text{Damage} = \text{plan to attack} \cdot (1 - R_{terrorists}) \cdot (1 - R_{public}) \cdot (1 - R_{private}) \quad \text{equation 2}$$

$$\Pr(R_{terrorists} = 1) = 1 - \Pr(R_{terrorists} = 0) = W_{public} \cdot p_{canceling} \quad \text{equation 3}$$

$$R_{public} = W_{public} \cdot S_{public} \quad \text{equation 4}$$

$$R_{private} = W_{private} \cdot S_{private} \quad \text{equation 5}$$

Whereby R_{public} and $R_{private}$ represent the response of public and private to warnings and are functions of warning issuance and sensitivity to warning. As $R_{terrorists}$, R_{public} and $R_{private}$ increase damage from a planned attack decreases. S_{public} and $S_{private}$ are public and private sensitivities to warnings.

Now, we can randomly create positive and negative events in the model consistent with the signal detection framework and examine effectiveness of issuing warnings through time. We assume $d' = 2$ and base rate = 5%, and later we discuss the sensitivity

of the results to these assumptions. Figure 1 shows an example of a base run of the model in a four-year time period.

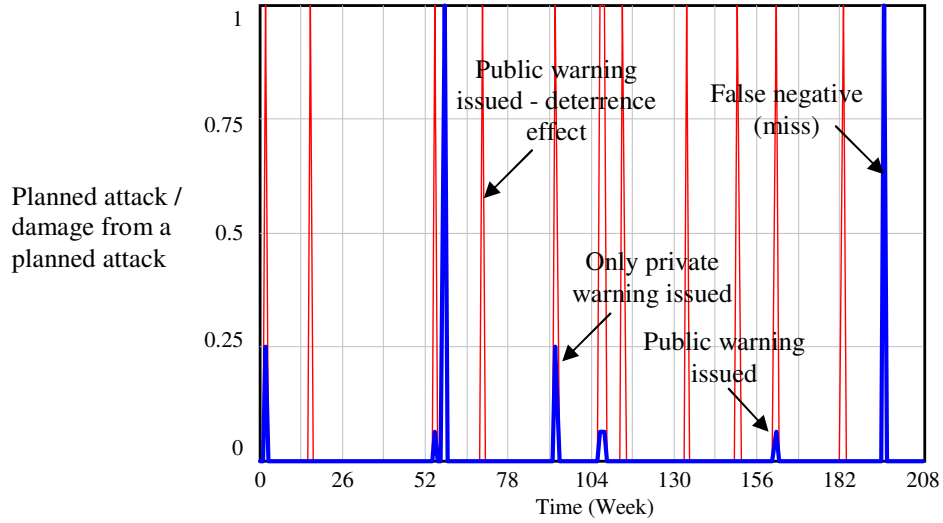


Figure 1. An example of a base run result for constant public and private sensitivity to warnings in four years - Darker lines show damage from a planned attack and lighter lines show planned attacks. The result shows two successful, two semi successful, three unsuccessful, and six canceled attacks; total damage of

2.75 units. ($S_{\text{public}} = S_{\text{private}} = 0.75$, $C_{\text{public}} = 1$, $C_{\text{private}} = 0.5$, $P_{\text{canceling}} = 0.5$)

2.2. Modeling Social sensitivity to warnings

Let's represent agents' memory of false alarms by a stock variable, Φ_i , which influences sensitivity to warning, S_i , as following:

$$\Delta\Phi_i = FP_{i.} \cdot (1 - \Phi_i) / \tau_1 - (TP_i + FN_i) \cdot \Phi_i / \tau_2 \quad \text{equation 6}$$

$$S_i = S_{\text{max}} - \alpha_i \cdot f_1(\Phi_i) \quad \text{equation 7}$$

In which $f_1(0) = 0$, $f_1(1) = S_{\text{max}} - S_{\text{min}}$, and $\dot{f}_1(\cdot) > 0$. α_i is the elasticity of agents to $f_1(\Phi)$.

We assume $f_1(x) = x \cdot (S_{\text{max}} - S_{\text{min}})$. Obviously, when $\alpha_i = 0$, $S_i = S_{\text{max}} = \text{constant}$. Changing α allows us to examine different assumptions about how people react to false alarms. τ_1

and τ_2 represent how fast agents update their memory when they receive false alarm or observe positive events respectively. As stated in equation 6, it is assumed that when $FP_i = 1$, Φ_i moves proportionally toward 1, and when either of TP_i or FN_i are equal to 1, Φ_i moves toward 0. Intuitively we can assume $\tau_1 > \tau_2$.

Figure 2 shows a base run of the simulation model for $\alpha_1 = \alpha_2 = 1$. In comparison to Figure 1 which uses the same random seeds, we see more damage under this scenario.

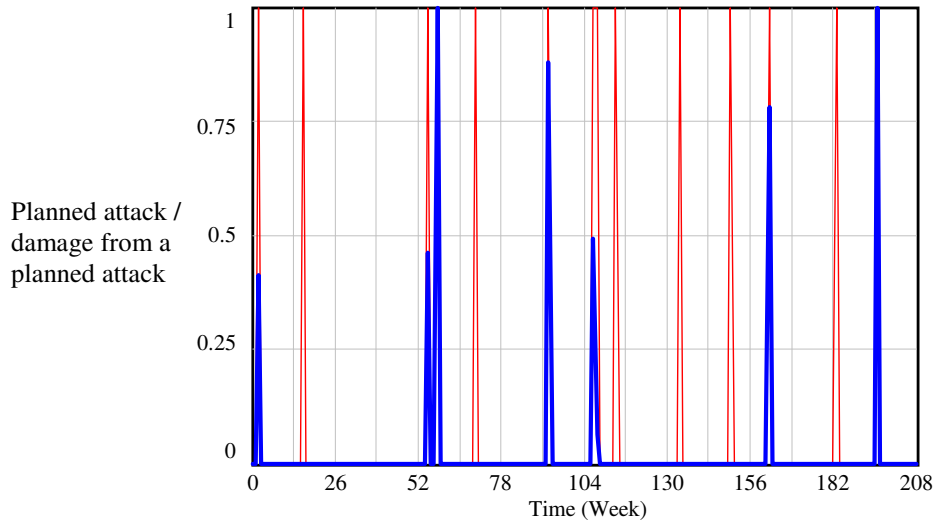


Figure 2. The base run result for changing public and private sensitivity to warnings in four years - Dark lines show damage from planned attack and lighter lines show planned attacks. The result shows total damage of 5.08 units. ($S_{\max} = 0.75$, $S_{\min} = 0.1$, $\alpha_1 = \alpha_2 = 1$, $\tau_1=4$ weeks, $\tau_2=1$ week, $C_{\text{public}} = 1$, $C_{\text{private}} = 0.5$, $P_{\text{canceling}} = 0.5$)

2.3. Modeling terrorists' reaction to social sensitivity

It maybe argued that it is too simplistic to assume that deterrence effect always exists even if public loses its sensitivity to warnings due to extensive false alarms. In fact, terrorists can update their perception about public sensitivity to warnings. In such a condition, if terrorists perceive lack of public sensitivity to warnings they may not cancel their plans to attack as frequently as they would if the public was responsive to warnings.

This hypothesis leads to a balancing loop in which when false alarms are issued (or perceived), public sensitivity (S_{public}) decreases, and that leads to a decline in terrorists perception of public sensitivity (\bar{S}_{public}) with some delay (Δt). As result, the chance of canceling a planned attack decreases, and more likely the society faces an attack. In order to formulate this loop we can use the following equations:

$$\Delta \bar{S}_{public,t} = (S_{public,t} - \bar{S}_{public,t}) / \Delta t \quad \text{equation 8}$$

$$P_{canceling} = P_{canceling,max} - \beta \cdot f_2(\bar{S}_{public,t}) \quad \text{equation 9}$$

In which $f_2(1) = 0$, $f_2(0) = P_{canceling,max} - P_{canceling,min}$, and $\dot{f}_2(\cdot) < 0$. For simplicity, we assume $f_2(x) = (1 - x^2) \cdot (P_{canceling,max} - P_{canceling,min})$. Figure 3 compares the base run simulation for different scenarios to examine the effect of terrorists' reaction to public sensitivity to warning. As we see total damage increases if terrorists update their perception of public sensitivity.

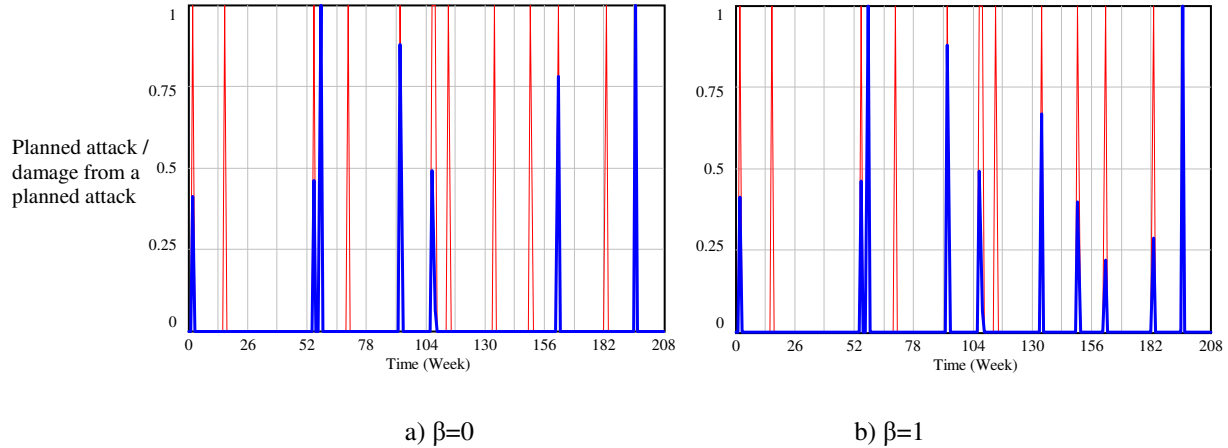


Figure 3. Comparing base run for two different conditions about how elastic terrorists are to change in public sensitivity - Dark lines show damage from planned attack and lighter lines show planned attacks. ($S_{max} = 0.75$,

$S_{min} = 0.1$, $\alpha_1 = \alpha_2 = 1$, $\tau_1=4$ weeks, $\tau_2=1$ week, $C_{public} = 1$, $C_{private} = 0.5$, $P_{canceling,max} = 0.5$, $P_{canceling,min} = 0$)

So far we examined dynamic base runs under different conditions just to get an idea of the dynamic trend of attacks and damages. In the next section we use the model to find optimal warning thresholds that minimize total damage in long-term under different scenarios.

3. Optimal Solution

We can conduct simulation runs for different random seeds and thresholds and find the thresholds that minimize the total cost of damage from attacks. Pinker (2007) assumes that the direct cost of issuing warnings (e.g., social stress) is negligible in comparison to the damage from an attack. In line with this assumption, we can assume that the goal of optimization is to find the decision thresholds that minimize damage from planned attacks over an extended period of time. For simplicity we assume that $C_{\text{private}} = C_{\text{public}} - 0.5$, so we can concentrate on optimizing one variable. Furthermore, different α and β represent different scenarios and we can test if a change in these scenarios has any influence on the optimal solution. Figure 4 shows average damage in a year (calculated in 20-year time interval) versus threshold for public warning under different scenarios.

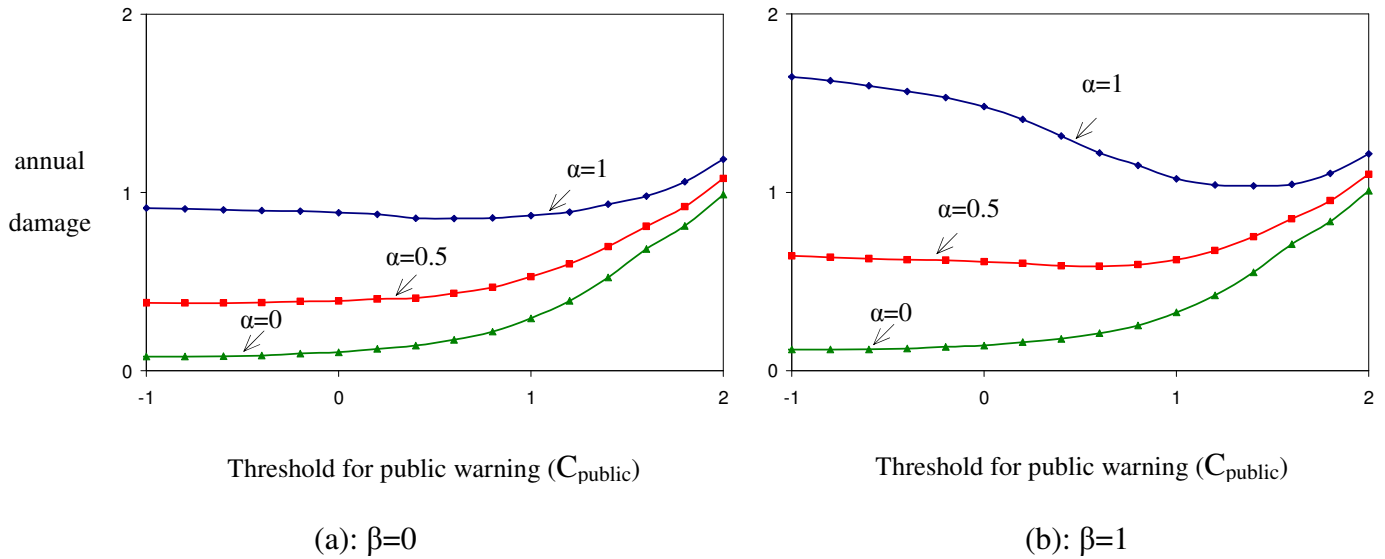


Figure 4 - average damage in a year versus threshold for public warning under different scenarios (different values of α and β).

Note: lower threshold causes more warning issuance.

In Figure 4, each graph is calculated based on more than 28,000 simulation runs. In Figure 4a, we examine only the effect of the first assumption (how sensitive public is to false alarms). As we see in small values for α (less sensitivity to false alarms) the model proposes extensive warning issuance and it shows that by increasing α warning issuance becomes less effective. Figure 4b includes the proposed balancing loop and assumes that terrorists can update their perception and thus deterrence effect can be influenced. In such a condition, as we see, the optimal solution changes, and as α increases the optimal threshold increases significantly. Figure 5 illustrates optimal public warning threshold under different scenarios (different α and β).

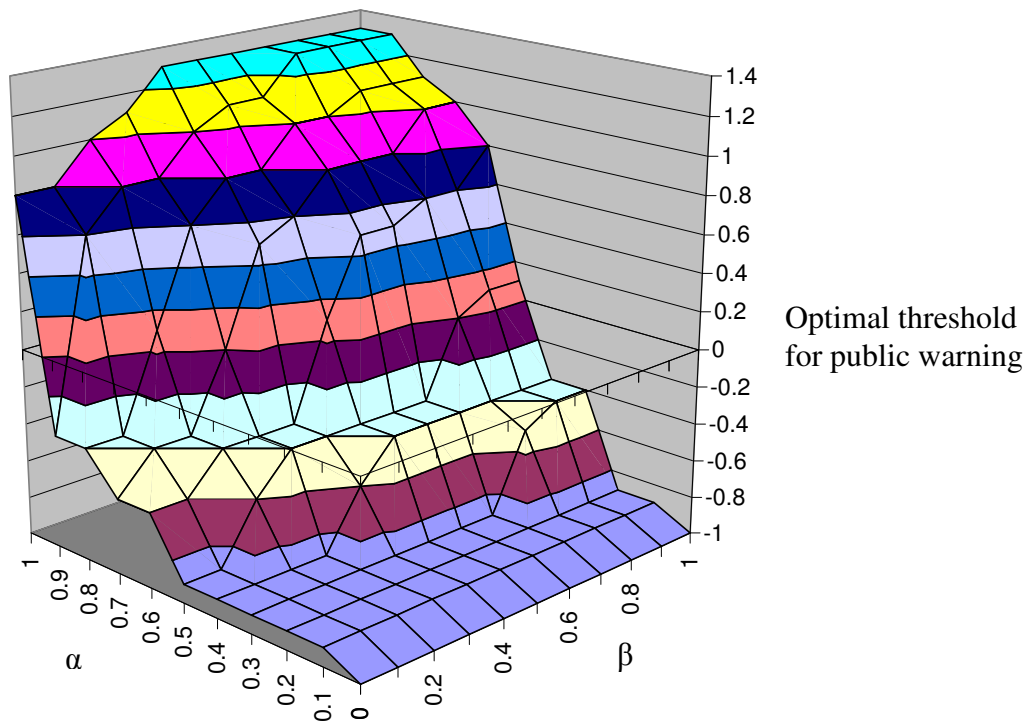


Figure 5 – Optimal threshold under different scenarios. Optimal threshold is sensitive to the scenarios: for higher α and β we get higher optimal thresholds, which recommends lower level of warning issuance.

As we see optimal threshold is sensitive to our assumptions about elasticity of public response to false alarms and terrorists reaction to loss of public sensitivity. In short it shows that the optimal solution can be biased if we underestimate the elasticity parameters introduced in this paper. As illustrated, for higher α and β we get higher optimal thresholds in comparison to lower values of α and β . Private warning threshold is assumed to be always 0.5 unit less than public warning threshold.

Finally we can examine the effect of intelligence system capabilities on optimal solutions for warning. As stated in section 2-1, the level of uncertainty in detecting signals from noise is determined by d' which can represent the intelligence system

capabilities in judging how severe an event looks like. Comparing the optimal solutions for different values of d' can be interesting. Interestingly, the model gives new insights about the effect of intelligence system on security policies, showing that there is an interactive effect between the intelligence system capabilities and how conservative the policymakers should be in issuing warnings.

Figure 5 compares annual damage under two scenarios of no and high elasticity to the mechanisms that we discussed ($\alpha = \beta = 0$ and $\alpha = \beta = 1$) for different values of d' . This figure has two main points. First, as intuitively we expect, the better the intelligence system the less successful attacks will a country face. In each of the conditions as d' increases the graph shifts downward. Second, in the left side figure ($\alpha = \beta = 0$) as the threshold decreases (more warning issuance) we get better results. It suggests that the optimal threshold is the lowest possible threshold, no matter what d' is. But in the right side figure we see that as d' increase the optimal threshold increases (for example compare graphs for $d'=3$ and $d'=2$. For the former one the optimal threshold is higher than for the latter one). Therefore, under stronger intelligence systems the bias induced by mechanisms described in this paper increases. This suggests that countries with better intelligence systems should consider the points from this model more seriously.

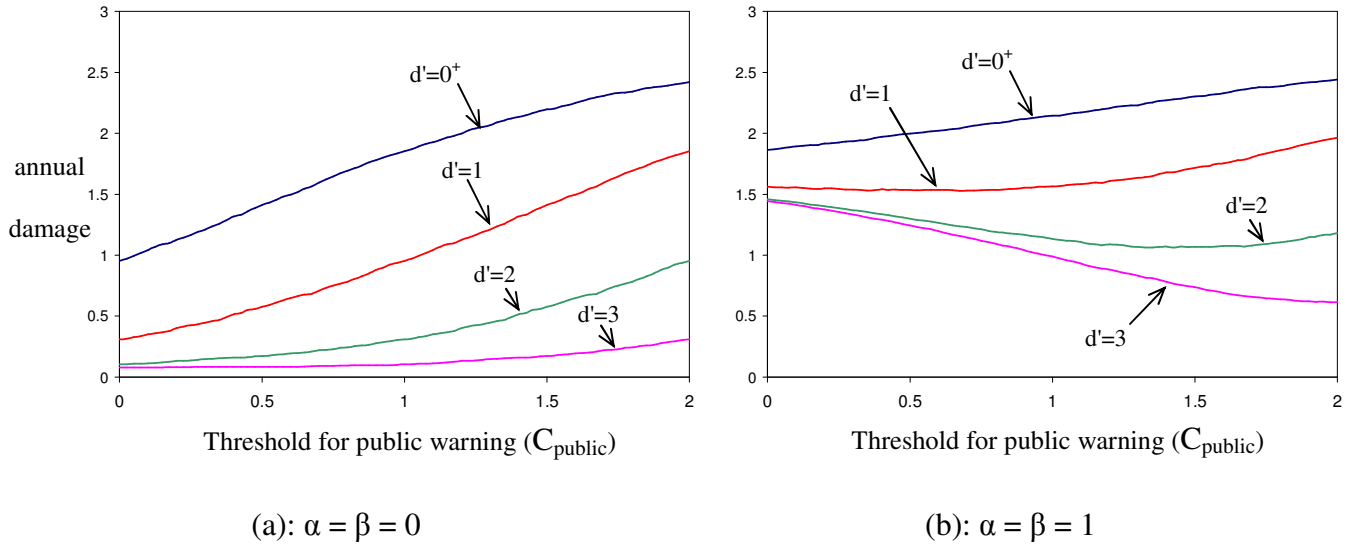


Figure 5 - average damage in a year versus threshold for public warning under different scenarios for intelligence system capabilities (different values of α and β).

4. Conclusion

We developed a simulation model based on signal detection framework that helped us to extend Pinker's (2007) framework of normative decision making on security. This research has two major contributions. First, the extended model helps us to find numerical solutions for the framework in a dynamic context and examine optimal private and public warning decisions.

Second, through the extended model we showed that optimal solutions are sensitive to two major assumptions: public sensitivity to false alarms and terrorists' perception of public sensitivity. Based on the assumption that as the public becomes more sensitive to false alarms, public responsiveness to alarms may decrease, we conclude that optimal solutions can be influenced by how sensitive public is to false alarms. Furthermore, if this sensitivity is perceived by terrorists while the magnitude of damage can change, the

optimal decision threshold changes. In summary, an underestimation of these effects can result in biases in optimal solutions. Furthermore, we argued that the magnitude of bias in countries with better intelligence systems can be higher.

There are several ways to extend this model. First, there can be an interactive effect between the base rate and decision thresholds. It can be expected that as in high base rates we will have less false alarms the effect of elasticity parameters on optimal solutions may change. Further, this model can also be used as a descriptive model to explain effectiveness of warning decisions in different contexts. The model can also be calibrated for different cultural contexts and examined to see how well it can describe warning policy performance.

References

- Forrester, J. W. 1971. Counterintuitive behavior of social systems. *Technology Review* 73 (3):52-68.
- Ghaffarzadegan, N. 2008, How a System Backfires: Dynamics of Redundancy Solution in Security, *Risk Analysis* 28(6): 1669 – 1687
- Green, D. M., & Swets, J. 1966, *Signal detection theory and psychophysics*. New York: Wiley.
- Macmillan, N. A., & Creelman, C. D. 1991. *Detection theory: A user's guide*. Cambridge: Cambridge University Press.
- Pinker, E. J. 2007. An Analysis of Short-term Responses to Threats of Terrorism. *Management Science* 53(6): 865-880
- Sagan, Scott D. (2004). The problem of redundancy problem, Why More Nuclear Security Forces May Produce Less Nuclear Security, *Risk Analysis*, Vol. 24, No. 4, 2004
- Swets, J. A., 1991, The science of high stakes decision making in an uncertain world (Transcript of a Science and Public Policy Seminar). Washington, D.C.: Federation of Behavioral, Psychological and cognitive sciences.