*Defence and Peace Economics*, 2011
Vol. 22(1), February, pp. 43–61

# SECRECY AND DECEPTION AT EQUILIBRIUM, WITH APPLICATIONS TO ANTI-TERRORISM RESOURCE ALLOCATION

JUN ZHUANG[a,*] AND VICKI M. BIER[b]

[a]*Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, 403 Bell Hall, Buffalo, NY 14260, USA;* [b]*Department of Industrial and Systems Engineering, University of Wisconsin-Madison; 3234 Mechanical Engineering Building, 1513 University Avenue, Madison, WI 53706, USA*

*(Received July 17 2009 in final form)*

In this work, we develop a game-theoretic model for whether and how a first mover should disclose her resource allocation. Our model allows us to explore whether the first mover should disclose correct information about her resource allocation, incorrect information, or no information. Although we study secrecy and deception specifically in the homeland-security context where the first mover is assumed to be the defender, our work can also provide insights in other contexts, such as business competition.

*Keywords:* Secrecy and deception; Truthful disclosure; Homeland security; Resource allocation; Game theory

## 1. INTRODUCTION

Both natural disasters and terrorism can cause huge casualties and economic losses. Moreover, both types of risks may require large investments beforehand to reduce the associated loss of life and economic impact. However, unlike natural disasters, attackers are intelligent and adaptive (Bier, 2005; Zhuang and Bier, 2007). Therefore, unlike investments in protection from natural disasters, which are usually disclosed to the public, anti-terrorism investments are not always disclosed. Understanding when and how such investment information should be disclosed becomes a big challenging issue for governments facing smart terrorists.

While it is relatively easy to find examples of secrecy and deception in international relationships and the military arena, it is difficult to find specific examples of government secrecy and deception in homeland security resource allocation, in part because these phenomena are often classified. In general, the use of government secrecy and/or deception generates criticism as being undemocratic (see Rourke, 1961; Galnoor, 1977; Cohen, 1990), and interfering with accountability (see for example Schneier, 2000; Clark *et al*., 1997; Rozell, 1994; Maskin and Tirole, 2004). There are also ethical concerns regarding government uses of secrecy and deception. Nevertheless, government secrecy and deception with

*Corresponding author. Email: jzhuang@buffalo.edu

ISSN 1024-2694 print: ISSN 1476-8267 online © 2011 Taylor & Francis
DOI: 10.1080/10242694.2010.491668

regard to resource allocation have occurred in the context of homeland security (see for example Wise, 1969). One example is the use of fake security cameras and guards to deter possible attacks.

In the literature, disclosure is often found to be preferable to secrecy. Levy (2007) notes that the benefits of truthful disclosure include 'enhanced accountability, enhanced predictability, and the provision of expert information to the economy.' Recent game-theoretic research has also indicated that publicizing defensive information instead of keeping it secret may help to deter attacks (see for example Sandler and Arce, 2003; Bier *et al*., 2007), or at least allow the defender to use her first-mover advantage[1] to 'guide' attackers towards less valuable or damaging targets (Zhuang and Bier, 2007). Similarly, Edmonds (2006) has argued that classifying too much information could in principle hurt national security. In other words, secrecy or deception might hurt information sharing between first responders or others responsible for security, and therefore might directly decrease the effectiveness of actual defenses. Although this type of disclosure is somewhat different from disclosure of defensive resource allocations (which is the topic of interest in this paper), it appears that at least under some circumstances, there can be merits to releasing defensive information.

In practice, however, security-related information such as defensive resource allocations is often kept secret. For example, the detailed allocation of onboard air marshals is usually kept secret (although the total number might be known at least approximately) in order to achieve attack deterrence. Similarly, as long as the installation of Lojack vehicle recovery on any given car is kept secret, Ayres and Levitt (1998) shows that information regarding market share of Lojack can help deter potential thefts. Finally, in the Iraq war, the US has largely been unable to announce 'good news' like the rebuilding of schools or hospitals, since doing so has often invited immediate attacks; disclosure of such good news unfortunately makes the newly rebuilt targets more attractive to attackers.

Defenders might also have incentives to deceive by either overstating or understating their defenses, to deter or disinterest potential attackers, respectively. We recognize that many types of secrecy and deception have of course been investigated using a variety of methods in the military (Joint Chiefs of Staff, 1996), psychology (DePaulo *et al*., 2003), and computer science (Swire, 2001, 2004), as well as economics and political science (as discussed below), but few of these studies focus specifically on disclosure of resource allocations.

### 1.1. Game Theoretical Modeling of Secrecy

Secrecy has been modeled as simultaneous play in game theory (see for example Zhuang and Bier, 2007), since, in a simultaneous game, each player moves without knowing the moves chosen by the other players. Note that this does not actually require both players to make their decisions at the same time – the players can be viewed as being engaged in a simultaneous game as long as neither party knows the other's decision at the time it makes its own decision. Probably the most similar work to ours, Brown *et al*. (2005), studies secrecy in a zero-sum attacker-defender game in the context of ballistic missile deployment, but fails to include the potential for the attacker to endogenously update his beliefs. In particular, in Brown *et al*. (2005), the attacker is assumed to be unaware even of the defender options when the defender chooses secrecy, while a typical endogenous model usually assumes only that the attacker is unaware of the specific choice made by the defender.

One possible reason for preferring secrecy to disclosure is when moving first confers no first-mover advantage over simultaneous play, which can easily occur when the action space

[1] For a general discussion of first-mover advantage in the economic literature, see for example Lieberman & Montgomery (1988, 1998).

is discrete. Simple examples include the symmetric zero-sum games of Matching Pennies and Rock/Paper/Scissors, in which no player wants to reveal his action before the other player commits to an action, since doing so would guarantee a loss. Instead, in these games, players would prefer to mix their strategies and move simultaneously. Thus, rational players would choose a symmetric strategy of 1/2 heads 1/2 tails in the game of Matching Pennies, and a symmetric strategy of 1/3 rock 1/3 paper 1/3 scissors in the game of Rock/Paper/Scissors. Dighe *et al*. (2009) apply a similar idea in the context of a two-target, attacker-defender game. Their model essentially compares four possible defender strategies: (1) defending both targets; (2) defending neither target; (3) defending one target with disclosure; and (4) defending one target without disclosing which target is defended. They argue that the fourth strategy is always preferred to the third (in other words, that ‘partial secrecy’ is better than full disclosure when only one target is defended), and in particular may allow the defender to deter an attack at less cost than defending both targets. We discuss the relationship between our work and that of Dighe *et al*. (2009) more thoroughly in Section 3.3.

## 1.2. Deception in the Literature

While the definition of secrecy is relatively straightforward, many kinds of deception have been discussed in the literature. Some researchers model deception as sending noisy or imperfect signals to mislead one’s opponents. For instance, Hendricks and McAfee (2006) and Oliveros (2005) use the Normandy invasion as an example to argue that the first mover (the Allies) optimally allocated resources to targets that they did not intend to attack, in order to mislead the Germans about their true landing place by sending noisy signals. Similarly, Hespanha *et al*. (2000) and Brown *et al*. (2005) define deception in a zero-sum attacker-defender game as occurring when the defender discloses only a subset of the defenses, in an attempt to route the attacks to heavily-defended locations.

In the context of voting, deception is sometimes defined as choosing an action that goes against one’s preferences. For example, Brams (1985) defines deception as ‘a player’s false announcement of his preferences to induce the other player to choose a strategy favorable to the deceiver.’ Similarly, Brams and Zagare (1977) define deception as ‘voting not directly in accordance with one’s preferences.’ (Note that this type of deception can occur at equilibrium only in games with more than two voters.)

In all of the examples of deception discussed above, the players reveal their actions (at least partially), but may choose actions that mislead other players about their type, and thus their future intentions. By contrast, we define deception as disclosing a signal (in the domain of the action space) that differs from the chosen (hidden) action. More specifically, in the context of homeland security, we interpret deception as disclosing a different level of defensive investment than what is actually implemented. Thus, we model secrecy and/or disclosure as a signal sent by the defender, while the true defense is treated as a (possibly) hidden action, since the actual level of defensive investment may not be directly observable by the attacker in our context. In principle, it may also be possible to model deception as a hidden action rather than a signal. However, in our view, deception requires not only a hidden action (the true level of defense), but also a deceptive disclosure. So, it seems appropriate to us to view the deceptive disclosure as a signal. Likewise, it may at first seem strange to view secrecy as a special case of a signal, but since secrecy in our model is in general more costly than truthful disclosure, it seems appropriate to view secrecy (i.e. a null signal) as a special case of a costly signal.[2] We

[2] The case in which the actual defense is guaranteed to be observed by the attacker can be obtained as a degenerate case of our model, by setting the deception and secrecy costs large enough that the defender will always choose truthful disclosure in any equilibrium.

also propose an equilibrium concept based on perfect Bayesian Nash equilibrium (PBE) to address this issue, in a fully endogenous signaling game with incomplete information (as defined by Spence, 1973) and hidden defender actions.

To our knowledge, this model is novel in the literature. We believe that explicitly modeling deception as a signal differing from a hidden action is a useful approach because, in reality, defenders often do have hidden actions available to them, while in traditional signaling games (see for example Crawford and Sobel, 1982; Cho and Kreps, 1987), signalers (first movers) do not have hidden actions. Moreover, in our model, the player utilities depend on the first player's actions, while in signaling models the player utilities depend only on the first player's type and signals, and the second player's actions (see for example Banks and Sobel, 1987; Cho and Kreps, 1987). Crawford (2003) defines deception similarly to our definition here, but in his context, such deception is found to occur only when there are some 'boundedly rational' (exogenous) players in the system because signals are assumed to be costless (cheap talk).

### 1.3. Other Related Literature

There is substantial economics literature on principal–agent or more generally mechanism-design problems, which address how the first mover (principal) can provide incentives (usually by contract) to the second mover (agent), to ensure that the second mover chooses the preferred action. This literature usually allows hidden actions and/or private information on the part of the second mover (Doepke and Townsend, 2006), and may further address the issue of information disclosure by the second mover (Prat, 2005), but not the first mover. By contrast, in Section 3 we argue that hidden actions and private information on the part of the second mover do not change the defender's preference for truthful disclosure in our model.

Another body of economics literature focuses on revelation of private information, addressing how individual players might either truthfully or deceptively disclose their private information or attributes (instead of player actions). For example, Gal-Or (1987) analyzes the effects of secrecy and/or truthful disclosure of information about consumer demand; Zhu (2004) analyzes whether suppliers should share cost data with their competitors; and Li (2002) analyzes whether retailers should disclose their uncertainties about costs and customer demand to manufacturers. Greenberg (1982) proposes a model for the effects of deception on the subjective probabilities of the second mover in a game (but fails to analyze how those subjective beliefs could be endogenously updated).

In the security context, Yetman (2004) shows that a discriminatory screening policy based on observed attributes (e.g. racial profiling) can be more efficient than a non-discriminatory one. Basuchoudhary and Razzolini (2006) study a similar attacker signaling game in which the security authority determines whether a passenger is a security threat based on observable attributes (e.g., race, country of origin) disclosed by the attacker, but conclude that a separating equilibrium usually will not exist; i.e., it would generally be sub-optimal to check only people with certain attributes.

Finally, cheap-talk games (see for example Farrell and Rabin, 1996) concern the effects of communication between players in situations where the players' utilities do not depend directly on the signal. Such costless talks are generally not credible, but may be useful in coordinating the actions of the players in cases with multiple equilibria. However, this is fundamentally different from our model, because the cost of signaling is critical to achieving credible and effective secrecy and deception. Moreover, there is good reason to believe that the cost of implementing effective secrecy or deception is likely to be non-negligible.

### 1.4. Outline of this Paper

The next section puts forth a basic model of secrecy and deception in the case of incomplete information, and defines what we mean by an equilibrium solution in this context. Section 3 provides a general proposition showing that the defender has a first-mover advantage in the special case with no private defender information. In Section 4, we show (by means of numerical examples) that secrecy and deception can sometimes be strictly preferred to truthful disclosure in games with defender private information. Sections 5 and 6 then give some future research directions, and conclude this paper.

## 2. MODEL FORMULATION FOR GAMES WITH INCOMPLETE INFORMATION

We start with a model of incomplete information, in which the attacker and/or the defender have some private information. In particular, we assume that the attacker and/or the defender are of particular ‘types,’ which are known to themselves but not to the other player. Although the realizations of their types are not observable to others, the ex ante probability distributions of their types are assumed to be common knowledge to both the attacker and the defender. We will let nature make the initial (zero-stage) move, randomly drawing the players’ types from the ex ante probability distribution. This kind of game is sometimes called a Bayesian game (see Fudenberg and Tirole, 1991: Chapter 8.2; Mas-Colel *et al*., 1995: Chapter 8.E).

In general, we suspect that the defender’s preference among the three options (secrecy, truthful disclosure, and deceptive disclosure) will depend both on her own type, and on the perceived (distribution of the) attacker’s type. From the economics literature about signaling and screening, if the main uncertainty in the game comes from the attacker’s lack of knowledge about the defender’s type, then the defender may wish to use signaling to mislead the attacker in a sequential game. By contrast, if the main source of uncertainty comes from the defender’s lack of knowledge about the attacker’s type, then the defender may use screening to help identify the attacker (although gaining more information about the attacker may not always benefit the defender, especially in a one-stage game). In Section 3, we will show that defender uncertainty about the attacker’s private information will never result in the defender preferring secrecy or deception to truthful disclosure. Terrorist signaling in the case where the defender is uncertain about the attacker’s private information has been studied by Lapan and Sandler (1993), but is beyond the scope of this paper.

### 2.1. Notation and Problem Formulation

We define the parameters of our model as follows:

- *A* and *D*: Attacker (signal receiver) and defender (signal sender), respectively.
- $a \in A$ and $d \in D$: Attacker effort and actual defensive investment, respectively, and *A* and *D* are assumed to be discrete and finite with cardinalities $|A|$ and $|D|$, respectively.
- $\Delta_k$: The ($k$–1)-dimensional simplex; i.e.,

$$\Delta_k \equiv \{\delta = (\delta(1),\ldots,\delta(k)) : \delta(i) \geq 0 \; \forall i = 1,\ldots,k \text{ and} \sum_{i=1}^{k} \delta(\mathrm{i}) = 1\} \tag{1}$$

- $\Delta_{|A|}$ and $\Delta_{|D|}$: the mixed extensions of *A* and *D*, respectively (see Mas-Colel *et al*., 1995: 232) defined by equation (1), for $k = |A|$ and $k = |D|$, respectively.

- $\Delta_S \equiv \{S\} \cup \Delta_{|D|}$: The set of possible signals that the defender can send to the attacker, consisting of secrecy ($\{S\}$) and disclosure of any possible defensive investment strategy in $\Delta_{|D|}$. Note also that sometimes we use quotation marks ‘ ’ to distinguish a signal from the actual defensive investment, to avoid confusion.
- $\theta_A \in \Theta_A$ and $\theta_D \in \Theta_D$: Random variables of attacker and defender types, respectively, where $\Theta_A$ and $\Theta_D$ are the sets of feasible types for $\theta_A$ and $\theta_D$, respectively. We assume that $\theta_A$ and $\theta_D$ are independent of each other, and $\Theta_A$ and $\Theta_D$ are finite with cardinalities $|\Theta_A|$ and $|\Theta_D|$, respectively. It is important to note that there do not actually need to exist multiple types of attackers and defenders. We require only that attacker subjectively believe that the defender can be of multiple types (with those beliefs described by a probability distribution) and vice versa, and that such beliefs are common knowledge.
- $p_A : \Theta_A \to [0,1]$ and $p_D : \Theta_D \to [0,1]$: The defender’s ex ante probability distribution functions for $\theta_A$, and the attacker’s ex ante probability distribution function for $\theta_D$, respectively. In this model, we assume that the probability distributions $p_A(\theta_A)$ and $p_D(\theta_D)$ are common knowledge to both the attacker and the defender.
- $\sigma_A : \Delta_S \times \Theta_A \to \Delta_{|A|}$: The *action rule* for the attacker. That is, $\sigma_A(a|s,\theta_A)$ is the probability that an attacker of type $\theta_A$ would choose pure strategy $a$ when he receives signal $s$.
- $(\sigma_D \times s)$: $\Theta_D \to \Delta_{|D|} \times \Delta_S$: The *action rule* and *signaling rule* for the defender, respectively. That is, the defenders of type $\theta_D$ would choose pure strategy $d$ with probability $\sigma_D(d|\theta_D)$ and choose a signal $s$.
- $\mu(\theta_D \mid s)$: $\Delta_{|\Theta_D|} \to [0,1]$: The attacker’s ex post probability distribution function for $\theta_D$ after observing the signal $s \in \Delta_s$.
- $u_A(a,d,\theta_A,\theta_D)$: $A \times D \times \Theta_A \times \Theta_D \to \Re$ and $u_D(a,d,s,\theta_A,\theta_D)$: $A \times D \times \Delta_s \times \Theta_A \times \Theta_D \to \Re$: The utility functions for the attacker and the defender, respectively, for pure strategies $a \in A$ and $d \in D$.[3] For notational convenience, for fixed $a \in A$ and $d \in D$, we extend the defender utility function $u_D$ on the action space $A \times D$ to $\hat{u}_D$ on the action space $\Delta_{|A|} \times D$, to $\tilde{u}_D$ on the action space $A \times \Delta_{|D|}$, and to $U_D$ on the action space $\Delta_{|A|} \times \Delta_{|D|}$, respectively, by taking expected values:

$$\hat{u}_D(\sigma_A, d, s, \theta_A, \theta_D) = \sum_{a \in A} u_D(a, d, s, \theta_A, \theta_D)\sigma_A(a \,|\, s, \theta_A) \tag{2}$$

$$\tilde{u}_D(a, \sigma_D, s, \theta_A, \theta_D) = \sum_{d \in D} u_D(a, d, s, \theta_A, \theta_D)\sigma_D(d \,|\, \theta_D) \tag{3}$$

$$U_D(\sigma_A, \sigma_D, s, \theta_A, \theta_D) = \sum_{a \in A}\sum_{d \in D} u_D(a, d, s, \theta_A, \theta_D)\sigma_D(d \,|\, \theta_D)\sigma_A(a \,|\, s, \theta_A) \tag{4}$$

Similarly, for fixed $\theta_A \in \Theta_A$ and $\theta_D \in \Theta_D$, we extend the attacker utility function $u_A$ on the action space $A \times D$ to $U_A$ on the action space $\Delta_{|A|} \times \Delta_{|D|}$ by taking expected values:

$$U_A(\sigma_A, \sigma_D, \theta_A, \theta_D) = \sum_{a \in A}\sum_{d \in D} u_A(a, d, \theta_A, \theta_D)\sigma_D(d \,|\, \theta_D)\sigma_A(a \,|\, s, \theta_A) \tag{5}$$

In the numerical examples provided in Section 4, some additional notations are used:

- $P(a,d) : A \times D \to [0,1]$: The contest success function; i.e., the probability of a successful attack when the attacker chooses attack effort $a$ and the defender chooses defensive investment $d$.

[3] Note that the defender utility can in principle also reflect the cost of implementing the signal $s$ (i.e., the cost of implementing secrecy, truthful disclosure, or deceptive disclosure).

- $g(a,\theta_A)$: $A \times \Theta_A \rightarrow \mathbb{R}$: The cost to an attacker of type $\theta_A$ of choosing the attack effort *a*. We also assume $g(0,\theta_A) = 0 \; \forall \; \theta_A \in \Theta_A$; i.e., there is no cost to any type of attackers if he chooses not to attack.
- $h(d,s,\theta_D)$: $D \times \Delta_S \times \Theta_D \rightarrow \mathbb{R}$: The cost to a defender of type $\theta_D$ of choosing the signal *s* and defensive investment *d*.

### 2.2. Sequence of Actions in Extensive Form

Figure 1 shows our game in extensive form. The decision process is as follows: first, nature chooses the types of the attacker and defender ($\theta_A$ and $\theta_D$), according to the probability distributions $p_A(\theta_A)$ and $p_D(\theta_D)$. (Recall that the realization of the random variable $\theta_A$ is observable only to the attacker, and the realization of $\theta_D$ is observable only to the defender.) Second, a defender of type $\theta_D$ chooses a (possibly mixed) strategy $\sigma_D(\theta_D)$ and a signal $s(\theta_D)$. Finally, an attacker of type $\theta_A$ responds to the observed signal *s* by choosing a (possibly mixed) attacker effort $\sigma_A(s,\theta_A)$, leading to attacker and defender total utilities given by $U_A[\sigma_A(s,\theta_A), \sigma_D(\theta_D),\theta_A,\theta_D]$ and $U_D[\sigma_A(s,\theta_A), \sigma_D(\theta_D),s(\theta_D),\theta_A,\theta_D]$, respectively. The attacker's response $\sigma_A(s,\theta_A)$ is determined endogenously in this model.

### 2.3. Assumptions

While we believe that our game is reasonably general, we do restrict our attention to the case of a single centralized defender and a single attacker (e.g., the US versus al Qaeda). Thus, we do not address how a single defender should allocate defenses against multiple different attackers (such as network administrator versus computer hackers). Likewise, we do not address the case of decentralized defenders (such as multiple countries, companies, or government agencies), in which case defenses by one agent could deflect attacks onto other agents, creating economic externalities (see Enders and Sandler, 1993; Kunreuther and Heal, 2003; Bier *et al*., 2007; and Zhuang *et al*., 2007).

We view secrecy as a special case of a signal. Unlike hidden actions, signals are assumed to be always observable by the attacker. We allow the defender to disclose mixed defenses (either truthfully or deceptively) – recognizing, for example, that disclosure of a random strategy for allocating air marshals to planes, or police patrols to buildings, could achieve better attack deterrence than disclosure of the actual allocation. However, we do not allow the defender to randomize between secrecy and disclosure. In particular, the sequential nature of

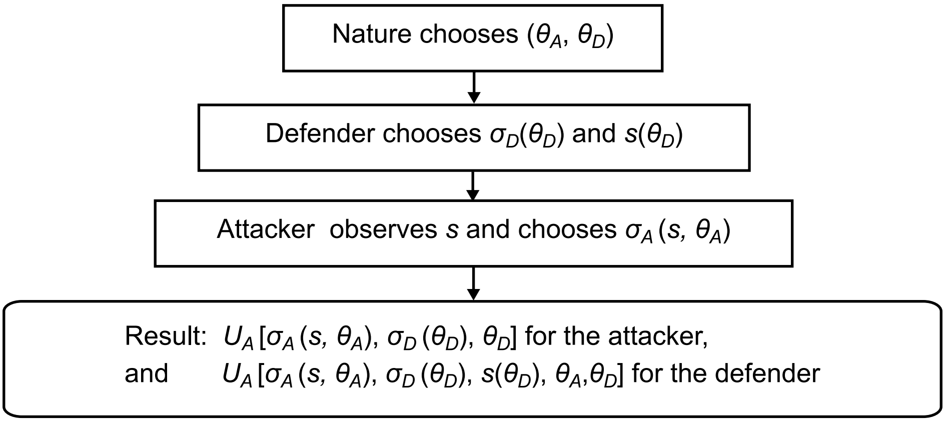
Nature chooses $(\theta_A, \theta_D)$
Defender chooses $\sigma_D(\theta_D)$ and $s(\theta_D)$
Attacker observes $s$ and chooses $\sigma_A(s, \theta_A)$
Result: $U_A[\sigma_A(s, \theta_A), \sigma_D(\theta_D), \theta_D]$ for the attacker,
and $U_A[\sigma_A(s, \theta_A), \sigma_D(\theta_D), s(\theta_D), \theta_A,\theta_D]$ for the defender

FIGURE 1 Sequence of actions for defender and attacker with private information

play in our game means that the choice of secrecy versus disclosure would be observable by the attacker in any case (since the attacker would know whether the defender had made a disclosure before having to choose an attack strategy), so there would be no benefit of randomization.

We assume that the attacker knows that the target exists, even if the defender keeps her defensive investment secret. We recognize that in some cases, defenders may use secrecy to prevent their opponents from even knowing the existence of a target at all. For example, a government would have no reason to disclose the defenses of a top-secret nuclear facility in the desert, if the attacker did not even know that the target existed. Nevertheless, many targets of interest with regard to homeland security (e.g., the Sears Tower, the Pentagon, the Golden Gate Bridge) are already well known to potential attackers. For targets like these, we expect that defenders may have incentives to deceive by overstating their defenses, but might not be able to disguise these targets as being of low value by understating their defenses. By contrast, for targets whose values may not be well known to attackers (such as some types of commercial databases), we speculate that defenders may in some cases have incentives to understate their defenses, to avoid ‘tipping off’ potential attackers to the values of those targets by disclosing large defensive investments.

Although we allow the attacker to update his knowledge about defender type through observing the signal sent by the defender, we do not allow other types of attacker observations and detections (such as spying or probing attacks), for reasons of simplicity. Finally, we assume that both the attacker and the defender are fully rational, and would like to maximize their utilities in the game specified in our model. Again, this is a restrictive assumption, but this may be the best assumption for the defender to make in the absence of information about the actual attacker goals and behavior.

### 2.4. Equilibrium

Analogous to perfect Bayesian equilibrium for signaling games without first-mover hidden actions (see Fudenberg and Tirole, 1991: Chapter 8.2), we define the equilibrium of our game with first-mover hidden actions as follows.

**Definition 1.** We call the collection of the attacker and defender action rules ($\sigma^*_A$ and $\sigma^*_D$), the defender signaling rule $s$*, and the attacker posterior belief $\mu$*, an *equilibrium* if equations (6)–(8) below are satisfied:

$$\sigma^*_D(\theta_D), s^*(\theta_D) \in \arg\max_{\sigma_D \in \Delta_{|D|}, s\in\Delta_s} \sum_{\theta_A \in \Theta_A} U_D\left[\sigma^*_A(s,\theta_A), \sigma_D, s, \theta_A, \theta_D\right] p_A(\theta_A) \quad \forall \theta_D \in \Theta_D \tag{6}$$

$$\sigma^*_A(s,\theta_A) \in \arg\max_{\sigma_A \in \Delta_{|A|}} \sum_{\theta_D \in \Theta_D} U_A\left[\sigma_A, \theta^*_D(\theta_D), \theta_A, \theta_D\right] \mu^*(\theta_D | s) \quad \forall s \in \Delta_s, \theta_A \in \Theta_A \tag{7}$$

and

$$\forall\, s \in \Delta_s, \quad \mu^*(\theta_D | s) = \frac{I_{\{S^*(\theta_D)=S\}} P_D(\theta_D)}{\sum_{\theta_{D'} \in \Theta_D} I_{\{S^*(\theta_{D'})=S\}} p_D(\theta_{D'})} \forall\, \theta_D \in \Theta_D$$
$$\text{if} \sum_{\theta_{D'} \in \Theta_D} I_{\{S^*(\theta_{D'})=S} P_D(\theta_{D'}) > 0 \tag{8}$$

where $I_{\{\cdot\}}$ is the indicator function.

In words, in order for $(\sigma^*_A, \sigma^*_D, s^*, \mu^*)$ to be an equilibrium, equation (6) requires that the defender's action rule $\sigma^*_D$ and signaling rule $s^*$ must maximize her expected utility given the attacker's response $\sigma^*_A$; equation (7) requires that the attacker's strategy $\sigma^*_A$ must maximize his expected utility given his posterior belief $\mu^*$ and the defensive investment $\sigma^*_D$ of those defender types; and equation (8) requires that the attacker's posterior belief $\mu^*$ is rational given the defender's signaling rule $s^*$, in the sense that the attacker uses Bayes' rule to determine $\mu^*(\theta_D|s)$. Note that if for some signal $s$, we have $I_{\{s^*(\theta_D)=s\}}p_D(\theta_D) = 0 \; \forall \; \theta_D \in \Theta_D$ (so that is an out-of-equilibrium signal or off-equilibrium path signal; see Cho and Kreps, 1987), then equation (8) does not restrict $\mu^*(\theta_D|s)$.[4] However, in order for the defender to optimize her actions and signals in equation (6), a well-defined equilibrium concept must specify the consequences for the defenders of choosing out-of-equilibrium signals. Therefore, equation (7) must hold for *all* signals $s \in \Delta_s$.

Based on the equilibrium concept in Definition 1, we define truthful disclosure, secrecy, and deception as follows.

**Definition 2.** In an equilibrium $(\sigma^*_A, \sigma^*_D, s^*, \mu^*)$, we say that a defender of type $\theta_D \in \Theta_D$ chooses

1. *truthful disclosure* if and only if $s^*(\theta_D) = \sigma^*_D(\theta_D)$;
2. *secrecy* if and only if $s^*(\theta_D) = \{S\}$; and
3. *deceptive disclosure* if and only if $s^*(\theta_D) \neq \{S\}$ and $s^*(\theta_D) \neq \sigma^*_D(\theta_D)$.

## 3. SPECIAL CASE – DEFENDER DOES NOT KNOWTHE ATTACKER TYPE

In this section, we consider the case where only the attacker has private information, while the defender has none. In particular, we show in Proposition 1 that in this case, the defender will always prefer truthful disclosure as long as the cost of implementing truthful disclosure is weakly lower than the cost of deception or secrecy. Note, by this way, that even though this section deals with cases of attacker private information, the results in this section also apply to the case of complete information as a special case. In Section 4, we will consider the case where the defender has private information.

### 3.1. Degenerate Equilibria

We model games with only private attacker information by setting the set of possible defender types $\Theta_D$ in the model of Section 2 to be a singleton, so that there is effectively no private defender information. In that case, equations (6)–(8) in Definition 1 reduce to equations (9) and (10) below,[5] yielding an equilibrium that consists of two action rules ($\sigma^*_A$ and $\sigma^*_D$) and one signaling rule $s^*$:

$$\sigma^*_D, s^* \in \underset{\sigma_D \in \Delta_{|D|}, s \in \Delta_s}{\arg\max} \sum_{\theta_A \in \Theta_A} U_D\left[\sigma^*_A(s,\theta_A), \sigma_D, s, \theta_A\right] P_A(\theta_A) \tag{9}$$

[4] Several bodies of literature provide refinements on equilibrium concepts for out-of-equilibrium signals. These include sequential equilibrium (Kreps and Wilson, 1982), divine equilibrium (Banks and Sobel, 1987), and stable equilibrium (Cho and Kreps, 1987). However, we do not further address this issue in our work, for reasons of simplicity.

[5] Note that we no longer need to update the attacker's belief about the defender type $\theta_D$, since there is no private information in this game. Therefore, equation (8) in Definition 1 is not needed. For simplicity, we also remove $\theta_D$ from equations (9) and (10), because it is a constant and common knowledge in this case.

$$\sigma^*_A(s,\theta_A) \in \underset{\sigma_A \in \Delta_{|A|}}{\arg\max}\; U_A\left[\sigma_A, \sigma^*_D, \theta_A\right] \quad \forall\, s \in \Delta_s, \theta_A \in \Theta_A \tag{10}$$

Note that the attacker response $\sigma^*_A$ in equation (10) above depends on the equilibrium signal $s^*$ only through its relationship to the defender action $\sigma^*_D$. In other words, different signals $s^*$ will affect the attacker's best response only if they are associated with equilibria involving different defender strategies $\sigma^*_D$. This shows that when the defender has no private information, defender signaling has no direct impact on the attacker response (because there is no attacker uncertainty), and therefore no effect on the defender's equilibrium utilities (except possibly through any direct costs).

### 3.2. Truthful Disclosure

Two factors determine the defender's preferences among the various possible signals (truthful disclosure, secrecy, and deception). The first is the direct or exogenous cost of implementing each type of signal. In other words, when the costs of secrecy and deception are sufficiently high, then truthful disclosure will of course be preferable. The second factor is the indirect effect of defender signals on (endogenously determined) attacker behavior.

In this paper, we assume that the cost of implementing truthful disclosure is always lower than the costs of secrecy and deception (since defenders must usually spend some extra effort to keep their actions secret, or to deceive their opponents). Thus, secrecy and deception will be preferred only because of their endogenous effects on attacker behavior (if at all). Proposition 1 below indicates that when secrecy and deception are costly compared with truthful disclosure (as assumed here), then they will never be strictly preferred by defenders in games without private defender information.

**Proposition 1.** In games with no private defender information, if the cost of implementing truthful disclosure is weakly lower than the costs of secrecy and deception (i.e., for any given attacker action $a$, attacker type $\theta_A$, and defense strategy $\sigma_D$, sending a truthful signal '$\sigma_D$' is no more costly for the defender than sending other signals),

$$'\sigma_D' \in \underset{s \in \Delta_s}{\arg\max}\, \tilde{u}_D(a, \sigma_D, s, \theta_A) \quad \forall a \in A, \theta_A \in \Theta_A, \sigma_D \in \Delta_{|D|} \tag{11}$$

then for any possible equilibrium $(\sigma^*_A, \sigma^*_D, s^*)$ in which the defender chooses secrecy or deception (i.e., $s^* \neq \sigma^*_D$), there must exist another equilibrium such that truthful disclosure yields the same defender expected utility.

**Proof.** For the proof see the Appendix.

Proposition 1 states that if truthful disclosure is the lowest-cost strategy for the defender, then it will also result in the highest defender utility in the case where only the attacker has private information. However, this is not always the case when the defenders have private information, as shown in Section 4.

### 3.3. Relationship to Previous Work

Proposition 1 suggests that there will be no deception by the defender when only the attacker has private information. However, deception by terrorists may well occur at equilibrium when the defender is uncertain about the attacker's private information (see for example Lapan and Sandler, 1993; Arce and Sandler, 2007).

Zhuang and Bier (2007) provide a result that can be viewed as a special case of Proposition 1 above. In particular, that work assumes complete information (a degenerate case of incomplete information), does not allow for the possibility of deception (or in other words assumes a prohibitively high cost of deception), and assumes that secrecy and truthful disclosure are equally costly. Therefore, the assumptions in Zhuang and Bier (2007) satisfy equation (11), and their finding that disclosure is preferred to secrecy is essentially a special case of Proposition 1.

Proposition 1 does not contradict the result that partial secrecy (disclosure of the total defensive investment, but not how that investment is allocated) is preferred to disclosure in Dighe *et al*. (2009). In particular, the model of Dighe *et al*. (2009) finds mixed strategies at equilibrium, but does not explicitly consider disclosure of the mixing probabilities as a defender option. However, since the attacker in the game of Dighe *et al*. (2009) can infer the mixing probabilities at equilibrium, truthful disclosure of those probabilities (as in Proposition 1) is unlikely to affect the outcome of the game, and therefore is likely to yield the same payoff as partial secrecy. (Similarly, we would argue that disclosure of the mixing probabilities is likely to yield the same payoffs as secrecy in the games of Matching Pennies and Rock/Paper/Scissors.)

## 4. SPECIAL CASE: DEFENDER KNOWS THE ATTACKER TYPES

Section 3 above addresses the case where only the attacker has private information, and shows that under reasonable conditions, the defender will prefer truthful disclosure to secrecy or deception. In this section, by contrast, we consider the case where only the defender has private information, while the attacker does not. We will show that, in this case, there exist equilibria in which some defender types strictly prefer secrecy or deception to truthful disclosure.

### 4.1. Degenerate Equilibria

In a game in which only the defender has private information, we can let $\Theta_A$ be a singleton. In this case, the three requirements for an equilibrium solution in Definition 1 reduce to Equations (12)–(14) below:[6]

$$\sigma_D^*(\theta_D), s^*(\theta_D) \in \underset{\sigma_D \in \Delta_{|D|},\, s \in \Delta_s}{\arg\max}\; U_D\left[\sigma_A^*(s), \sigma_D, s, \theta_D\right] \quad \forall\, \theta_D \in \Theta_D \tag{12}$$

$$\sigma_A^*(s) \in \underset{\sigma_A \in \Delta_{|A|}}{\arg\max} \sum_{\theta_D \in \Theta_D} U_A\left[\sigma_A, \sigma_D^*(\theta_D), \theta_D\right] \mu^*(\theta_D | s) \quad \forall\, s \in \Delta_s \tag{13}$$

and

$$\forall s \in \Delta_s,$$
$$\mu^*(\theta_D | s) = \frac{I_{\{s^*(\theta_D)=s\}} p_D(\theta_D)}{\sum_{\theta_{D'} \in \Theta_D} I_{\{s^*(\theta_{D'})=s\}} p_D(\theta_{D'})} \forall\, \theta_D \in \Theta_D \text{ if } \sum_{\theta_{D'} \in \Theta_D} I_{\{s^*(\theta_{D'})=s\}} p_D(\theta_{D'}) > 0 \tag{14}$$

[6] As in Section 3, for simplicity, we remove $\theta_A$ from equations (13) and (14), because it is a constant and common knowledge.

where $I_{\{\cdot\}}$ is the indicator function. As discussed in Subsection 2.4, note that equation (13) must hold for all signals $s \in \Delta_s$.

## 4.2. Proposition on Secrecy and Deception

Recall that when equation (11) is satisfied (e.g., when the cost of implementing truthful disclosure is lower than the costs of secrecy and deception), Proposition 1 suggests that truthful disclosure is always preferred to secrecy or deception for the case of private attacker information. An analogous condition for the case of private defender information is[7]

$$\sigma_D^* \in \arg\max_{s \in \Delta_S} \tilde{u}_D(a, \sigma_D, s, \theta_D) \quad \forall a \in A, \sigma_D \in \Delta_{|D|} \tag{15}$$

When Equation (15) is satisfied for a defender of type $\theta_D$, then for any fixed attacker effort and defensive investment, the cost of implementing truthful disclosure is lower than the costs of secrecy and deception. If a defender of type $\theta_D$ does not satisfy equation (15), then it is trivial to find an equilibrium in which that defender type would prefer secrecy or deception for the simple reason that truthful disclosure is costly in that case. However, we have a stronger result, as given in the following proposition.

**Proposition 2.** In games with private defender information, defenders may still strictly prefer secrecy or deception at equilibrium even if they satisfy equation (15) (i.e., if truthful disclosure is less costly than secrecy or deception).

We prove Proposition 2 by means of numerical examples. In particular, Section 4.3 introduces a simple 2×3×2 signaling game for this purpose and provides an example in which both defender types satisfy equation (15), but some of them still strictly prefer secrecy or deception.

## 4.3. A Simple 2×3×2 Signaling Game

We illustrate Proposition 2 by setting up a simple 2×3×2 signaling game, in which we allow the defender to choose whether to defend (i.e., $D = \{D_y, D_n\}$), and which of three signals to send – '$D_y$' ('defending'), '$D_n$' ('not defending'), or $S$ (secrecy). The attacker then chooses whether to attack (i.e., $A = \{A_y, A_n\}$). We consider the case in which the attacker is uncertain about the defender's asset valuation and the costs. However, we expect that similar results would hold if the attacker were uncertain about some other critical model parameters, such as the probability of a successful attack (see for example Powell, 2007), or the costs of defender signals or defenses. These asset valuations could presumably be based on economic value, number of fatalities, symbolic importance, or some combination of these; see for example O'Hanlon *et al*. (2002). Similarly, the costs could be either the defender's actual costs or the opportunity cost of implementing deception or secrecy.

For simplicity, we assume in the rest of this paper that the cost of a mixed strategy is sufficiently high so that neither player would ever choose a mixed strategy, and the defender would never send a mixed signal. The utility function for the defender consists of two terms – the expected loss of the target value (if any), and the combined cost of the signal and defense, $h(d,s,v)$. Similarly, the utility function for the attacker consists of two terms – the target value gained (if any), and the cost of the attack (if any), where the cost of an attack is assumed not

[7] Recall that $\theta_D$ was suppressed in the function $\tilde{u}_D$ in equation (11) for simplicity. Similarly, we suppress A from the function $\tilde{u}_D$ in equation (15).

to depend on the target valuation. We also assume that the attacker's gains from an attack are exactly equal to the defender's losses. To summarize, the utility functions are:

$$u_A(a,d,v) = p(a,d)v - g(A_y)\, I_{\{a=A_y\}} \tag{16}$$

$$u_D(a,d,s,v) = -p(a,d)\,v - h(d,s,v) \tag{17}$$

For purposes of illustration, let the contest success function,[8] $P(a,d)$, be given by

$$P(a,d) = \begin{cases} 0 & \text{if } a = A_n \\ 0.5 & \text{if } a = A_y,\, d = D_y \\ 1 & \text{if } a = A_y,\, d = D_n \end{cases} \tag{18}$$

Table I summarizes the attacker and defender utilities for all possible combinations of actions and signal for this 2×3×2 game.

Suppose we have two possible defender types, $\Theta_D = \{\theta_1, \theta_2\}$, with probabilities $p_D(\theta_1) = p_D(\theta_2) = 0.5$. Let $g_A(A_n) = 0$; $g_A(A_y) = 4$. We consider two cases: first, when the costs for the two types of defenders are equal; and then relaxing that assumption to allow different costs for different defender types.

*Case A: Defender Costs Are Equal*

| | $g_D\,(d, s, \theta_1) = g_D\,(d, s, \theta_2)$ | | |
|---|---|---|---|
| | $s$ = '$D_n$' | $s$ = '$D_y$' | $s = S$ |
| $d = D_n$ | 0 | 2 | 1 |
| $d = D_y$ | 6 | 4 | 5 |

Let the costs for both defenders be given by

Note that the above defender costs ensure that secrecy and deception are more costly than truthful disclosure for both defender types. Allowing the ranges of asset values for the two-defender types to be $0 < v_1,\, v_2 \leq 20$, the equilibrium defender actions are given in Figure 2.

In particular, when both asset valuations $v_1$ and $v_2$ are high (the '·' area in Figure 2), then the attacker will choose to attack with certainty, and both defender types will choose defense and truthful disclosure (an example of a pooling equilibrium). When both asset valuations $v_1$ and $v_2$ are low (the '×' area in Figure 2), the attacker will choose not to attack with certainty; both defender types will choose not to defend, and will truthfully disclose their lack of defense (another pooling equilibrium). However, when one asset valuation is high and the other is relatively low (the '□' and '❍' areas in Figure 2), then the defenders with high and low valuations will choose defense and lack of defense, respectively, and will disclose their actions (an example of a separating equilibrium).

[8] For more general contest success functions, see for example Skaperdas (1996).

TABLE 1 Utilities for Attacker and Defender with Private Defender Valuation

| | $a = A_y$ | $a = A_n$ |
|---|---|---|
| $(d,s) = (D_y, \text{‘}D_y\text{’})$ | $\frac{v}{2}-g(A_y),-\frac{v}{2}-h(D_y,' D_{y'},\theta)$ | $0,-h(D_y,\text{‘}D_y\text{’},\theta)$ |
| $(D_y,\text{‘}D_n\text{’})$ | $\frac{v}{2}-g(A_y),-\frac{v}{2}-h(D_y,' D_{n'},\theta)$ | $0, -h(D_y, \text{‘}D_n\text{’},\theta)$ |
| $(D_y,S)$ | $\frac{v}{2}-g(A_y),-\frac{v}{2}-h(D_y,' S,\theta)$ | $0, -h(D_y,S,\theta)$ |
| $(D_n,\text{‘}D_y\text{’})$ | $v-g(A_y), -v-h(D_n,\text{‘}D_y\text{’},\theta)$ | $0,-h(D_n, \text{‘}D_y\text{’},\theta)$ |
| $(D_n, \text{‘}D_n\text{’})$ | $v-g(A_y), -v-h(D_n, \text{‘}D_n\text{’}, \theta$ | $0, -h(D_n, \text{‘}D_n\text{’}, \theta)$ |
| $(D_n,S)$ | $v-g(A_y), -v-h(D_n, S, \theta)$ | $0, -h(D_n, S, \theta)$ |

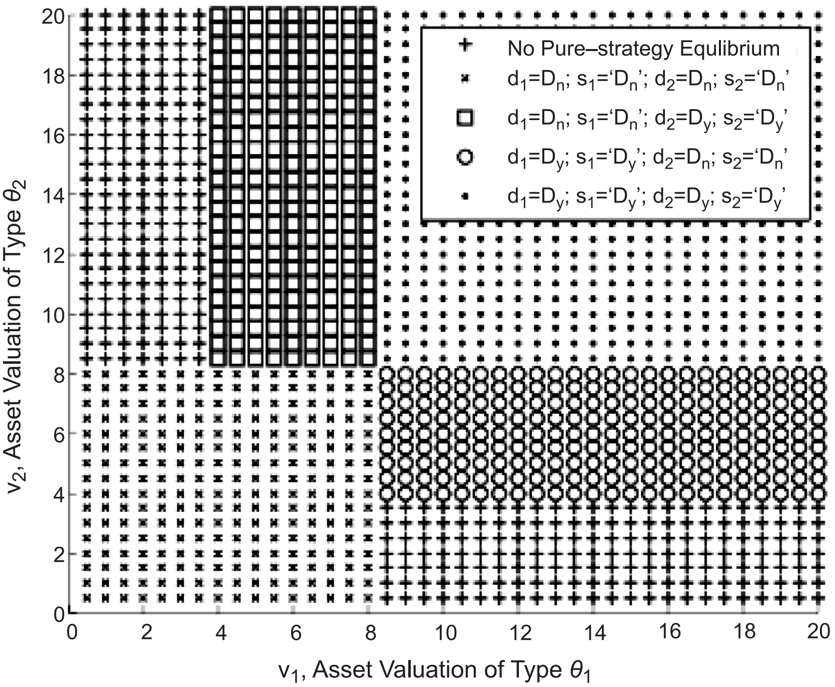

FIGURE 2 Equilibrium defender actions as a function of defender asset valuations when costs are the same

Finally, when one asset value is extremely high and the other is extremely low (the ‘+’ areas in Figure 2), then the low-value defender type will choose not to defend. However, the high-value defender in that case will prefer to defend when the attacker chooses to attack, and prefer not to defend when the attacker chooses not to attack. Therefore, this case does not have any pure-strategy equilibrium. A mixed-strategy equilibrium might well exist, but is beyond the scope of this paper.

### *Case B: Defender Costs Differ*

As noted above, we do not find secrecy and/or deception when the costs are equal. Therefore, we now revise the parameters of the model so that the defender of type $\theta_1$ has higher costs of deception and secrecy than the defender of type $\theta_2$:

| | $g_D\,(d, s,\theta_1)$ | | | $g_D\,(d, s, \theta_2)$ | | |
|---|---|---|---|---|---|---|
| | $s$ = ‘$D_n$’ | $s$ = ‘$D_y$’ | $s = S$ | $s$ = ‘$D_n$’ | $s$ = ‘$D_y$’ | $s = S$ |
| $d = D_n$ | 0 | 12 | 11 | 0 | 2 | 1 |
| $d = D_y$ | 16 | 4 | 15 | 6 | 4 | 5 |

Note that the above defender costs still ensure that secrecy and deception are more costly than truthful disclosure for both defender types. As before, we allow the ranges of asset values for the two-defender types to be $0 < v_1, v_2 \le 20$. The equilibrium defender actions are presented in Figure 3. (Note that the asymmetry in Figure 3 comes about because of the differences in the defender costs.)

In most regions of Figure 3, the equilibrium actions are the same as in Case A. However, for some combinations of asset valuations (the ‘□’ areas in Figure 3), the defender with lower deception and secrecy costs (type $\theta_2$) overstates her defenses to mimic the defender with higher costs, in order to achieve attack deterrence at low cost by free riding (an example of deception that results in a pooling equilibrium). When her asset valuation is relatively low, the defender of type $\theta_2$ may also choose secrecy (the ‘□’ areas in Figure 3, an example of a separating equilibrium). It may at first seem surprising that the defender of type $\theta_2$ invests in secrecy when her asset value is so low as not to interest the attacker in any case; however, this is done to prevent the defender of type $\theta_1$ from masquerading as being of type $\theta_2$ (and thereby attracting attacks against both types of defenders).

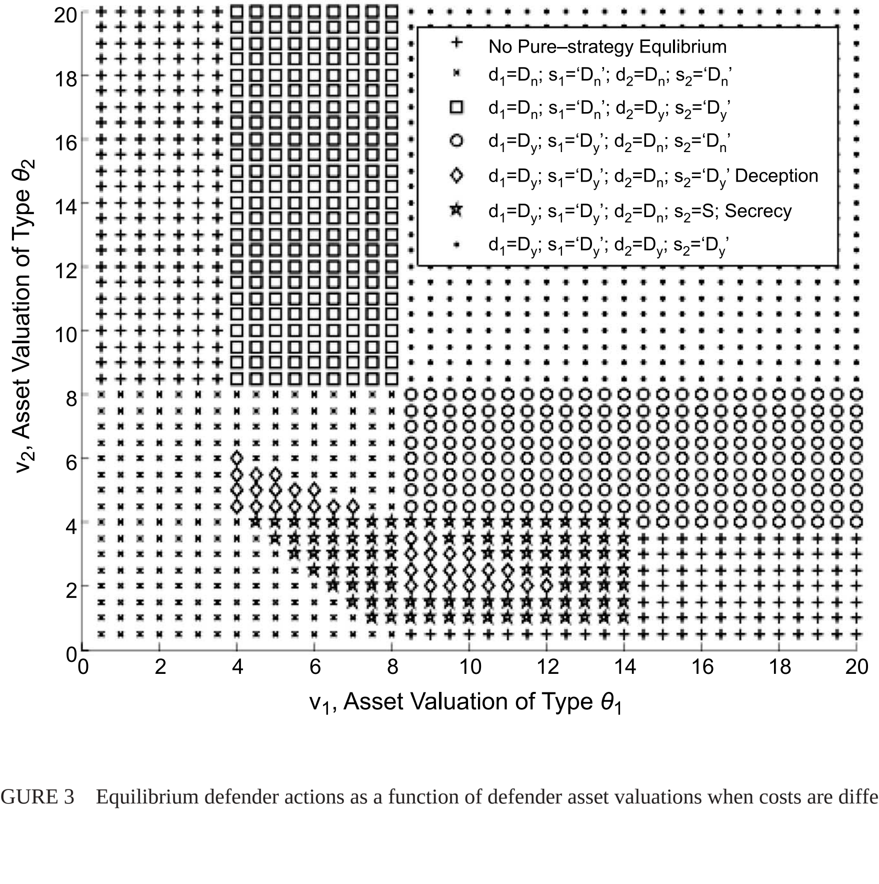

FIGURE 3 Equilibrium defender actions as a function of defender asset valuations when costs are different

In general, it has proven difficult to provide comparative statics for this model analytically. However, Figures 2 and 3 provide some indication of the types of results that can be observed as target valuations change. Moreover, comparing Figure 2 (where both defenders have the same costs of secrecy and deception) with Figure 3 (where one type has higher secrecy and deception costs) illustrates the effects of changes in signaling costs. In particular, when both types have low secrecy and deception costs (as in Figure 2), secrecy and deception do not appear at equilibrium, at least in pure strategies (although this type of behavior might conceivably occur in mixed strategies). However, when one type of defender has higher secrecy and deception costs, the defender type that is advantaged by lower secrecy and deception costs can sometimes exploit that advantage by free riding on the defensive investments of the other defender type (the '□' areas), or deflecting attacks to the other defender type (the '□' areas).

## 5. FUTURE RESEARCH DIRECTIONS

We have shown here that secrecy and/or deception can be equilibrium strategies for a defender. In particular, we find that the low-value defender may either choose secrecy, or else overstate her defense. However, it remains to clarify the conditions under which these strategies are preferred. For example, we speculate that when the attacker condition is continuous, or could play a mixed-strategy, we may also find understatement of defenses by high-value defenders. Moreover, once it is known that secrecy or deception may be optimal strategies, this opens up the question of whether it may sometimes be optimal for the defender to allocate defensive resources to targets that are not among those most attractive to attackers (unlike the recommendations in Bier *et al*., 2007); e.g., if those targets can be defended more cost-effectively than the most attractive targets.

In addition, our model addresses only a single-stage game, while in the real world, attackers and defenders interact repeatedly over time either through successive attacks (as in Israel, for example), or through successive attacker attempts to 'probe' the actual defenses of a system by launching small attacks and observing the results (e.g., in the context of computer security). In that case, secrecy and deception might be preferred by relatively short-sighted defenders (i.e., defenders with high discount rates), since the defenders' actual actions could be detected sooner or later. To address this situation, we plan to analyze a general *N*-stage game between a single attacker and a single defender, in order to more realistically model secrecy and deception in homeland security resource allocation (see Zhuang *et al*., 2010, for an example). This should allow us to model phenomena such as:

1. attacker learning over time (e.g., through repeated probes);
2. defender reputation effects (in which, for example, deception might be desirable in the short term, but lead to loss of credibility in the long term);
3. changes in the level of defense over time (e.g., increasing defensive investment in response to increased risk, or decreasing investment due to complacency in the absence of recent attacks); and
4. changing attacker and defender 'technologies' (e.g., implementation of new attack strategies).
5. Changing the resources available for defense and attack to model income effects over time.

Obviously, such models will almost certainly not be analytically tractable if they reflect anything approaching realistic levels of complexity. However, they should be readily solvable using dynamic programming.

## 6. SUMMARY

Our study represents one of the first rigorous attempts to explicitly compare all three possible disclosure strategies – truthful disclosure, secrecy, and deception – in the context of a contest against an intelligent and adaptable adversary. In particular, we show that all three signals (truthful disclosure, secrecy, or deception) are possible at equilibrium. Although we model deception and secrecy in the context of counter-terrorism, our results could also be applied in other areas, including military behavior, contracts, and business competition. For example, while corporate secrecy might seem to be generally advisable, in some cases disclosure of large investments might deter possible competitors from entering a new market. Thus, models similar to ours could be used to decide whether and how a company should disclose its strategic capital investments (e.g., in new-product development or marketing) to its competitors.

## ACKNOWLEDGMENTS

This research was supported by the United States Department of Homeland Security through the Center for Risk and Economic Analysis of Terrorism Events (CREATE) under grant number 2007-ST-061-000001. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security. We also thank Professor Larry Samuelson (Yale University), Professor Todd Sandler (University of Texas at Dallas) and Mr Stephen Campbell (Tufts University), and two anonymous referees for their helpful comments. Any errors are ours, of course.

### *References*

Arce, D.G. and Sandler, T. (2007) Terrorist signalling and the value of intelligence. *British Journal of Political Science* **37** 573–586.

Ayres, I. and Levitt, S. (1998) Measuring the positive externalities from unobservable victim precaution: an empirical analysis of lojack. *The Quarterly Journal of Economics* **113**(1) 43–77.

Banks, J. and Sobel, J. (1987) Equilibrium selection in signaling games. *Econometrica* **55**(3) 647–661.

Basuchoudhary, A. and Razzolini, L. (2006) Hiding in plain sight – using signals to detect terrorists. *Public Choice* **128**(1–2) 245–255.

Bier, V. (2005) Game-theoretic and reliability methods in counter-terrorism and security. In *Mathematical and Statistical Methods in Reliability,* Series on Quality, Reliability and Engineering Statistics, edited by A. Wilson, N. Limnios, S. Keller-McNulty and Y. Armijo. Singapore: World Scientific, pp. 17–28.

Bier, V., Oliveros, S. and Samuelson, L. (2007) Choosing what to protect. *Journal of Public Economic Theory* **9**(4) 563–587.

Brams, S. (1985) *Superpower Games: Applying Game Theory to Superpower Conflict.* New Haven, CT: Yale University Press.

Brams, S. and Zagare, F. (1977) Deception in simple voting games. *Social Science Research* **6** 257–272.

Brown, G., Carlyle, M., Diehl, D., Kline, J. and Wood, K. (2005) A two-sided optimization for theater ballistic missile defense. *Operations Research* **53**(5) 263–275.

Cho, I. and Kreps, D. (1987) Signaling games and stable equilibria. *The Quarterly Journal of Economics* **102**(2) 179–222.

Clark, G., Jonson, E. and Caldow, W. (Eds) (1997) *Accountability and Corruption: Public Sector Ethics.* St. Leonards, NSW, Australia: Allen & Unwin.

Cohen, S. (1990) *Government Secrecy in Democracies.* Cambridge, MA: Educators for Social Responsibility.

Crawford, V. (2003) Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review* **93**(1) 133–149.

Crawford, V. and Sobel, J. (1982) Strategic information transmission. *Econometrica* **50**(6) 1431–1451.

DePaulo, B., Wetzel, C., Sternglanz, R. and Wilson, M. (2003) Verbal and nonverbal dynamics of privacy, secrecy, and deceit. *Journal of Social Issues* **59**(2) 391–410.

Dighe, N., Zhuang, J. and Bier, V. (2009) Secrecy in defensive allocations as a strategy for achieving more cost-effective attacker deterrence. *International Journal of Performability Engineering* **5**(1) 31–43.

Doepke, M. and Townsend, R. (2006) Dynamic mechanism design with hidden income and hidden actions. *Journal of Economic Theory* **126**(1) 235–285.

Edmonds, S. (2006) Porter Goss' op-ed: 'ignoturn per ignotius'! Accessed January 2010. URL: http://www.truthout.org/article/sibel-edmonds-porter-gosss-op-ed-ignoturn-ignotius

Enders, W. and Sandler, T. (1993) The effectiveness of anti-terrorism policies: Vectorautoregression-intervention analysis. *American Political Science Review* **87**(4) 829–844.

Farrell, J. and Rabin, M. (1996) Cheap talk. *The Journal of Economic Perspectives* **10**(3) 103–118.

Fudenberg, D. and Tirole, J. (1991) *Game Theory.* Cambridge, MA: The MIT Press.

Gal-Or, E. (1987) First mover disadvantages with private information. *Review of Economic Studies* **54**(2) 279–292.

Galnoor, I. (1977) *Government Secrecy in Democracies.* New York, NY: Harper & Row.

Greenberg, I. (1982) The role of deception in decision theory. *Journal of Conflict Resolution* **26**(1) 139–156.

Hendricks, K. and McAfee, P. (2006) Feints. *Journal of Economics & Management Strategy* **15**(2) 431–456.

Hespanha, J., Ateskan, Y. and Kizilocak, H. (2000) Deception in non-cooperative games with partial information. In *Proceedings of the Second DARPA-JFACC Symposium on Advances in Enterprise Control.* URL: http://www.ece.ucsb.edu/~hespanha/published/deception.pdf

Joint Chiefs of Staff (1996) Joint doctrine for military deception. Joint Publication, 3-13.4, URL: http://www.c4i.org/jp3_13_4.pdf

Kreps, D. and Wilson, R. (1982) Sequential equilibria. *Econometrica* **50**(4) 863–894.

Kunreuther, H. and Heal, G. (2003) Interdependent security. *Journal of Risk and Uncertainty* **26** 231–249.

Lapan, H.E. and Sandler, T. (1993) Terrorism and signalling. *European Journal of Political Economy* **9**(3) 383–397.

Levy, G. (2007) Decision making procedures for committees of careerist experts. *American Economic Review, Papers and Proceedings* **97**(2) 306–310.

Li, L. (2002) Information sharing in a supply chain with horizontal competition. *Management Science* **48**(9) 1196–1212.

Lieberman, M. and Montgomery, D. (1988) First-mover advantages. *Strategic Management Journal* **9** 41–58.

Lieberman, M. and Montgomery, D. (1998) First-mover (dis)advantages: retrospective and link with the resource-based view. *Strategic Management Journal* **19** 1111–1125.

Mas-Colel, A., Whinston, M. and Green, J. (1995) *Microeconomic Theory.* New York: Oxford University Press.

Maskin, E. and Tirole, J. (2004) The politician and the judge: accountability in government. *American Economic Review* **94**(4) 1034–1054.

O'Hanlon, M., Orszag, P., Daalder, I., Destler, I., Gunter, D., Litan, R. and Steinberg, J. (2002) *Protecting the American Homeland: A Preliminary Analysis.* Washington, DC: Brookings Institution Press.

Oliveros, S. (2005) Equilibrium bluffs: a model of rational feints. Working paper, University of Wisconsin-Madison, Department of Economics.

Powell, R. (2007) Allocating defensive resources with private information about vulnerability. *The American Political Science Review* **101**(4) 799–809.

Prat, A. (2005) The wrong kind of transparency. *American Economic Review* **95**(3) 862–877.

Rourke, F. (1961) *Secrecy and Publicity: Dilemmas of Democracy.* Baltimore, MD: Johns Hopkins Press.

Rozell, M. (1994) *Executive Privilege: The Dilemma of Secrecy and Democratic Accountability.* Baltimore, MD: The Johns Hopkins University Press.

Sandler, T. and Arce, D.G. (2003) Terrorism and game theory. *Simulation & Gaming* **34** 319–337.

Schneier, B. (2000) *Secrets and Lies: Digital Security in a Networked World.* Hoboken, NJ: Wiley.

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

Spence, A. (1973). Job market signaling. *Quarterly Journal of Economics* **87**(3) 355–374.

Swire, P. (2001) What should be hidden and open in computer security: lessons from deception, the art of war, law, and economic theory. ArXiv Computer Science e-prints, (p. cs/0109089).

Swire, P. (2004) A model for when disclosure helps security: what is different about computer and network security? *Journal on Telecommunications and High Technology Law* **2** 1–38.

Wise, D. (1969) *The Politics of Lying: Government Deception, Secrecy, and Power.* New York: Random House.

Yetman, J. (2004) Suicidal terrorism and discriminatory screening: An efficiency-equity trade-off. *Defence & Peace Economics* **15**(3) 221–230.

Zhu, K. (2004) Information transparency of business-to-business electronic markets: a game-theoretic analysis. *Management Science* **50**(5) 670–685.

Zhuang, J. and Bier, V. (2007) Balancing terrorism and natural disasters – defensive strategy with endogenous attacker effort. *Operations Research* **55**(5) 976–991.

Zhuang, J., Bier, V. and Alagoz, O. (2010) Modeling secrecy and deception in a multiple-period attacker-defender signaling game. *European Journal of Operational Research* **203**(2) 409–418.

Zhuang, J., Bier, V. and Gupta, A. (2007) Subsidies in interdependent security with heterogeneous discount rates. *The Engineering Economist* **52**(1) 1–19.

## APPENDIX – PROOF OF PROPOSITION 1

Proposition 1 follows from the fact that the defender always has the option to choose and truthfully disclose $\sigma^*_D$, since the attacker response in equation (10) depends on the equilibrium signal $s^*$ only through its relationship to the defender action $\sigma^*_D$. In other words, signals $s^*$ will affect the attacker's best response only if they are associated with equilibria involving different defender strategies $\sigma^*_D$.

Specifically, if for some $s^*$ there exists an equilibrium $(\sigma^*_A, \sigma^*_D, s^*)$ that satisfies equation (10), then by the definition of an equilibrium, the corresponding collection $(\sigma^*_A, \sigma^*_D, {}'\sigma^*_D{}')$ involving truthful disclosure also satisfies equation (10). Moreover, equation (11) implies that

$$\tilde{u}_D(a, \sigma^*_D, {}'\sigma^*_D{}', \theta_A) \geq \tilde{u}_D(a, \sigma^*_D, s^*, \theta_A) \quad \forall a \in A, \theta_A \in \Theta_A$$

$$\Rightarrow \sum_{a \in A} \tilde{u}_D(a, \sigma^*_D, {}'\sigma^*_D{}', \theta_A)\, \sigma^*_A(a \mid s, \theta_A) \geq \sum_{a \in A} \tilde{u}_D(a, \sigma^*_D, s^*, \theta_A) \sigma^*_A(a \mid s, \theta_A) \quad \forall \theta_A \in \Theta_A$$

$$\Rightarrow U_D(\sigma^*_A, \sigma^*_D, {}'\sigma^*_D{}', \theta_A) \geq U_D(\sigma^*_A, \sigma^*_D, s^*, \theta_A) \quad \forall \theta_A \in \Theta_A$$

$$\Rightarrow \sum_{\theta_A \in \Theta_A} U_D(\sigma^*_A, \sigma^*_D, {}'\sigma^*_D{}', \theta_A)\, p_A(\theta_A) \geq \sum_{\theta_A \in \Theta_A} U_D(\sigma^*_A, \sigma^*_D, s^*, \theta_A) p_A(\theta_A)$$

Since by equation (9), the defender action rule $\sigma^*_D$ and the signaling rule $s^*$ on the right-hand side of the last line above are already maximal, the last inequality above must be an equality. Therefore, the collection $(\sigma^*_A, \sigma^*_D, {}'\sigma^*_D{}')$ satisfies equation (9), and is an equilibrium giving the same expected defender utility as $(\sigma^*_A, \sigma^*_D, s^*)$.