Attacking a Nuclear Facility with a Noisy Intelligence and Signal Disruption

Dov Biran Siyu Ma[1] * Yair Tauman[1,2] †

July 21, 2019

Abstract

The paper analyzes the interaction between two enemy nations – Player 1 (the weak nation) and Player 2 (the strong nation). 1 (he) wishes to develop a nuclear bomb and 2 (she) who employs an intelligence system, IS, aims to deter him by attacking 1. If 1 refuses to open his facilities for inspection, 2's IS will send a noisy signal indicating whether 1 builds a bomb (b) or not (nb). 1 has a disruptive technology (DT) which disrupts IS's signal with positive probability. Based on the signal sent by IS, 2 decides whether to attack 1 or not. The precision of IS is 2's private information while the quality of DT is common knowledge. The paper characterizes the unique perfect Bayesian equilibrium of the game and produces some surprising results. (1) Operating a better-quality DT lowers 1's payoff, therefore increases the probability 1 allows inspection and prevents a conflict. (2) Suppose 2 receives the signal b. If 1 estimates IS to be of high quality, 2 attacks 1 with probability 1. If, however, 1 estimates IS to be of low quality, 2 does not attack 1 with significant probability. (3) As expected, a more precise IS benefits 2, and for some parameters even benefits 1.

*[1] Adelson School of Entrepreneurship, Interdisciplinary Center (IDC), Herzliya
†[2] Economics Department, Stony Brook University

1 Introduction

The paper analyzes the interaction between two enemy countries. Player 1 is a “weak” country and Player 2 is a “strong” Country. Player 1 (he) wishes to develop a nuclear bomb or any weapon of mass destruction, and Player 2 (she) aims to deter him either by requiring 1 to open his facilities for inspection, or by attacking his facilities if 2 believes 1 builds a bomb.
Player 2 has an imperfect intelligence system (IS) which sends one of two signals: b to indicate “ there is a bomb” an nb to indicate “no bomb”. The precision of IS is α, namely, IS sends the right signal with probability α, $\frac{1}{2} < \alpha < 1$. Based on the signal of IS and on her belief on the action taken by 1, Player 2 decides whether to attack 1.
Player 1 can prevent 2 from attacking him by opening his facilities to reveal his innocence. However, allowing inspection is costly for him. It may reflect 1’s weakness in regime, make 1 loose face and may expose military details 1 wishes to conceal. Player 1 has a disruptive technology (DT), which with probability β makes IS completely obsolete (reduces α to $\frac{1}{2}$) or fails to disrupt IS with probability $1 - \beta$. He operates DT only if he builds a bomb (if he does not build a bomb, it is to his interest to have IS as accurate as possible). Unless 1 opens his facility for inspection, 2 does not know if DT is used (since she does not know whether 1 builds a bomb).
The cost of 1 allowing inspection and the precision of IS are private information of Player 1 and 2 respectively, but their probability distributions are commonly known. For simplicity we also assume β is common knowledge. This reflects on the “strong” Player 2 being informed about the technology of the “ weak” Player 1.
There are five possible outcomes of the two players’ interaction:
$(O, NA), (NB, NA), (NB, A), (B, NA) and (B, A)$, where O stands for opening the facilities for inspection, B for building a bomb, NB for not building a bomb, NA for not attacking 1 and A for attacking 1. Player 1 ranks the last four outcomes (from best to worst) as $(B, NA) \succ_1 (NB, NA) \succ_1 (NB, A) \succ_1 (B, A)$. Also $(NB, NA) \succ_1 (O, NA)$, but the exact ranking of (O, NA) depends on 1’s cost of being inspected. Player 2’s ranking (from best to worst) is $(O, NA) \sim_2 (NB, NA) \succ_2 (B, A) \succ_2 (NB, A) \succ_2 (B, NA)$. The sequence of events starts with 1 choosing either O, NB or B. If 1 chooses O, 2 does not attack and the game is over; otherwise, IS is sent out to detect whether 1 builds a bomb or not. If 1 chooses B he also operates DT, trying to disrupt IS’s signal. After receiving the signal sent by IS, 2 decides whether to attack 1. This together with cardinal utilities representing the above preference orders of the players define a game of incomplete information or simply, a Bayesian game. It is shown that the game has a unique Perfect Bayesian Equilibrium and it is characterized as follows. There exists a pair of thresholds $(\bar{c}, \bar{\alpha})$ s.t. Player 1 with inspection cost below $\bar{c}$ chooses not to build a bomb and open his facility for inspection. If the cost exceeds $\bar{c}$, he mixes the two strategies of building and not building a bomb. Depending on 1’s expectation on IS’s precision ($E(\alpha)$), the equilibrium belongs to one of the following two categories:
(i) Suppose Player 1 estimates IS to be of better quality $E(\alpha) \geq \bar{\alpha}$. Both players act

conservatively. Player 1, believing that 2 has a good IS, assigns a relatively low probability on building a bomb. If α exceeds a certain threshold $\hat{\alpha_1}$, Player 2 follows the signal (attack if the signal is b and not attack if the signal is nb). If α is below $\hat{\alpha_1}$, 2 does not rely on IS and attack irrespective of the signal, in attempt to avoid attacking the innocent opponent by mistake. Hence it is possible for Player 2 to miss a bomb builder even if IS sends the right signal.
(ii) Suppose Player 1 estimates IS to be of lesser quality, $E(\alpha) < \bar{\alpha}$. Both players then act aggressively. Player 1 builds a bomb with high probability. If α exceeds a certain threshold, Player 2 follows the signal. If α is below this threshold, Player 2 attacks 1 irrespective the signal to avoid missing a bomb builder. In particular, (1) suppose 2 receives the signal b. If 1 estimates IS to be of low quality, 2 attacks 1 with probability 1; if, however, 1 estimates IS to be of high quality, then with significant probability 2 refrains from attacking 1. (2) The tragedy that 1 does not build a bomb, IS sends the right signal but 2 still attacks 1 happens with positive probability.

A better disruption technology (higher β), surprisingly, lowers 1's payoff (if he refuses inspection) and increases the probability of 1 opening his facilities for inspection. A higher disruptive capability of DT stimulates Player 2 to act more aggressively, attacking 1 on a wider range of α and reducing his expected payoff. Consequently, 1 is better off not using DT if he can credibly commit to do so.
Not surprising, it is shown that 2's payoff is non-decreasing in the precision of IS, irrespective of 1's estimate over IS's quality. It is strictly increasing in α in the region where 2 in equilibrium follows the signal, and it does not depend on α in the region where 2 ignores the signal.
If the actual precision of IS is high, underestimating the quality of IS (α is high while $E(\alpha)$ is low) is costly for 1. However, if IS's precision is low, 1 may benefit from overestimating the quality of IS. High estimate may result in conservative behavior of 1 (where he builds a bomb with low probability) and in "not attack" as the best reply of 2. A consistent estimate (α and $E(\alpha)$ are both low) may result in aggressive behavior of 1 and in "attack" as the best reply of 2. The former is better for 1 than the latter.

2 Related Literature

The framework of this paper is similar to Jelnov, Tauman and Zeckhauser (2017) (JTZ1 hereafter). In JTZ1, Player 2 acts cautiously if her IS is relatively accurate and aggressively if it is less accurate. In the former case, after receiving the signal nb 2 for sure does not attack 1, but if 2 receives the signal b she mixes the two strategies: attack and not attack. In the latter case, 2 attacks 1 if she receives the signal b, and she mixes the two strategies if she receives the signal nb. In contrast, in this paper, 2's choice of acting conservatively or aggressively depends not only on the actual precision of IS, but also on Player 1's estimation of the quality of IS. Player 2, in our model, always plays a pure strategy, irrespective of the precision of IS.
JTZ1 is extended toward another direction in Jelnov, Tauman and Zeckhauser (2018) (JTZ2

hereafter), where an additional type of Player 1, a provocateur (P), is considered. Unlike JTZ1 and the current paper, where Player 1's primary goal, whether or not he has built the bomb, is to avoid an attack, a provocateur, by contrast, prefers to be attacked when he does not build a bomb. (An unjustified attack would bring support to 1 and blame to 2.)

We extends the above two papers in two ways: (1) The precision of IS in both JTZ1 and JTZ2 is commonly known, while we assume it to be Player 2's private information. (2) Unlike our model, in JTZ1 and JTZ2, Player 1 is unable to disrupt the signal. Having these two features, we show (1) even though Player 1 is able to reduce the reliability of the signal, it turns out to lower his payoff. He is best off not using his technology if he can commit to do so. (2) Player 1's estimation over IS's quality determines whether the two players both act aggressively or conservatively. An overestimation may actually benefit 1.

Our paper is also related to two prior works, one by Baliga and Sjöström (2008) (hereafter B&S), the other by Debs and Monteiro (2014) (hereafter D&M). The model of Baliga and Sjortrom, like us, deals with asymmetric information. In B&S, Player 1 (the weak nation) can be either normal or crazy, e.g., would give weapons to terrorists. He also differs in the expected costs of building a bomb, which we ignore. Player 2 can be a peaceful dove (never attack), aggressive hawk (always attack), or an opportunistic type who does not attack if she believes 1 is normal and has the bomb. By refusing to open his facilities, Player 1 can maintain strategic ambiguity. Thereby, 1 can still deter 2 from attacking him and avoid the cost of building a bomb. Their model, opposite to ours, focuses on deterrence of Player 2 from attacking; while we emphasize Player 2's incentive to wipe out the bomb if it does exist. Greater ambiguity in B&S makes Player 1's decision to build a bomb less likely, whereas in our model greater ambiguity — as reflected in a less reliable intelligence system (whether or not caused by 1's disruption) — never makes the build decision less likely. B&S also provides an insightful analysis of the possible roles for direct communication between the players.

Debs and Monteiro provide an insightful analysis of why the United States (Player 2) invaded Iraq (Player 1). They attribute that decision to the combination of Iraq's inability to commit not to develop nuclear weapons, and the United States' inability to definitively conclude that Iraq was not pursuing such an effort. Their model shows that when 2 has the capability to launch a preventive war, and has highly capable intelligence, Player 1 will refrain from making a power-shifting military investment. He understands that a preventive war will be the result, making it worse off than if it had never invested. However, with lower intelligence capabilities, 2 may launch a preventive war, as the United States did against Iraq, even though Iraq did not possess the bomb.

Our model allows an opportunity for 1 to open facilities for international inspection. Player 1's type, as reflected by his disutility of opening facilities, is assumed to be unknown to Player 2 and typically it is significant. In equilibrium, for some parameters of our model, with positive probability (i) 1 will not develop the bomb, (ii) he refuses inspection, and (iii) 2 launches a preventive war even if her intelligence is of relatively high quality and sends the correct signal.

Our model also relates to the literature on inspection games, where an inspector verifies whether an agent(s) adheres to specified rules. The agents therefore alter their activities in

ways that seek to conceal the true situation. Avenhaus et al. (2002) provides an extensive survey of this literature. Inspection games are similar to our game in their sequence of moves. First, an agent decides whether or not to adhere to the rules. If he chooses NOT, he will send the optimal noisy signal of his action. The inspector observes the signal and decides whether or not to audit him. The analogy to our game is obvious. The agent is like our Player 1, and the inspector is like our Player 2. One important difference between our model and a typical inspection game is that in the latter, by auditing the agent, the inspector can detect with certainty whether or not he adhered to the rules, before possibly taking tough measures against him. In our model Player 2's tough action of attacking Player 1's facility is taken under uncertainty since she can't detect with certainty what action 1 took. Another difference is that opening for inspection is a choice of player 1 and he may prefer not opening to opening even if it means getting attacked. This can't happen in inspection games. Finally, the signal in inspection games is strategically designed and sent by the agent, while in our game, the signal is not Player 1's decision variable but rather the outcome of a machine owned by 2. However, in our model, 1 can impact the distribution of the signal using his disruptive technology.
Our model relates broadly to a variety of models in the arms building, nuclear deterrence, and more generally to military strategy. See O'Neill (1994) for an extensive survey of this literature.

3 Model

A weak nation, Player 1 (he), wants to build a bomb, and a strong nation, Player 2 (she), wants to prevent him from building it by attacking his facilities if necessary.
Player 1 is asked to open his facility for inspection. If he does not possess a bomb, he can avoid a potential attack by opening his facilities to prove his innocence. Player 1 incurs a cost (dis-utility) $c > 0$ for allowing inspections (loosing face and exposing military details are costly). Player 1's pure strategies are: B (to build a bomb and not open), NB (not to build and not open) and NBO (not to build and open). In case Player 1 does not open, 2 can either attack, A, or not attack, NA.
The following table describes the payoffs of the players of the five possible outcomes:

2 1	NA	A
NB	$w_1, 1$	r_1, r_2
B	$1, 0$	$0, w_2$
O	$w_1 - c, 1$	

Figure 1: Payoff Table

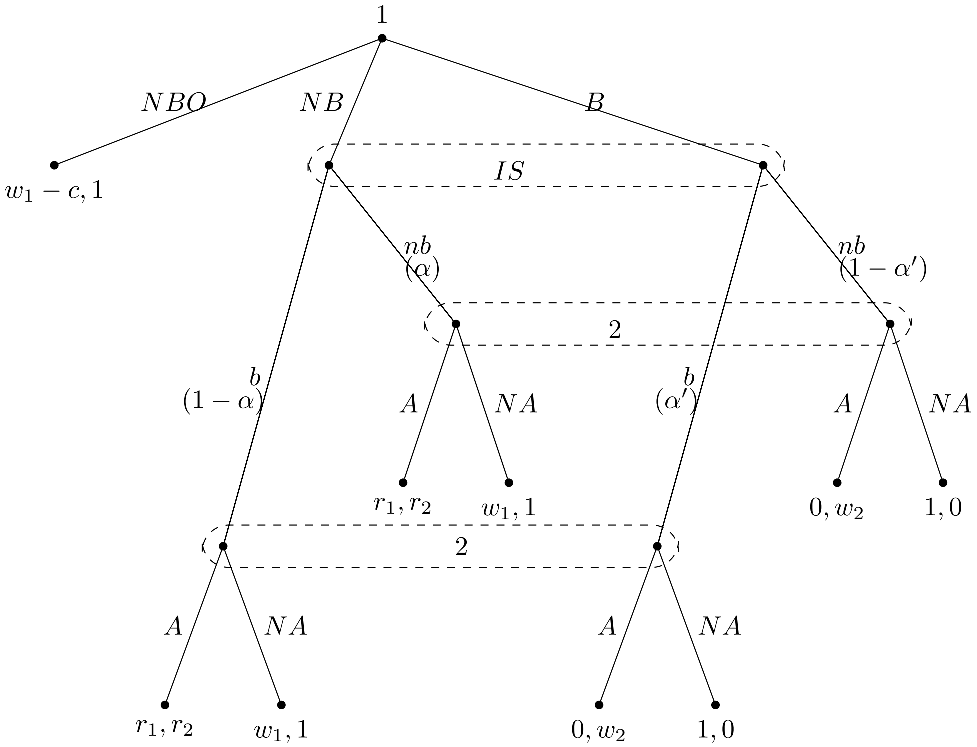
It is assumed that $c > 0,\ 0 < r_i < w_i < 1,\ i = 1, 2$. That is, 1 ranks the outcomes (from best to worst) as follows: (B,NA), (NB,NA) (NB,A) and (B,A). The outcome (O,NA) is ranked below (NB,NA); but its exact ranking depends on the value of c. Player 2 ranks the five possible outcomes (from best to worst) as follows: (NB,NA), (B,A), (NB,A) and (B,NA), and she is indifferent between (O,NA) and (NB,NA).

To determine whether or not to attack, Player 2 employs a noisy intelligence system (IS) to spy on her enemy and to detect if he builds a bomb. IS sends one of the two signals: b or nb, indicating whether 1 has a bomb or not, respectively. The precision of IS is α, $\frac{1}{2} < \alpha < 1$. Namely, if 1 chooses either B or NB, then with probability α, IS sends the signal b or nb respectively. IS with $\alpha = \frac{1}{2}$ is useless since the signal is sent completely randomly.

Player 1 has a disruptive technology, DT, which changes the distribution of the signal in the following way: with probability $\beta \in [0, 1)$, DT makes IS completely obsolete (the precision drops to $\frac{1}{2}$); with probability $1 - \beta$, DT fails to interfere the signal and IS's precision remains unchanged (namely, α). He uses DT only if he decides to build a bomb. (If 1 decides not to build a bomb, it is to his interest not to disrupt the signal.) So the precision of IS, if Player 1 chooses NB, is α (uninterrupted); if Player 1 chooses B the precision of IS is α', where $\alpha' = \frac{\beta}{2} + (1 - \beta)\alpha \leq \alpha$. Based on the signal she receives, Player 2 decides whether or not (or with what probability) to attack 1.

The set of pure strategies Player 2 is $\{(A, A), (A, NA), (NA, A), (NA, NA)\}$, where (X, Y) stands for playing X if the signal is b and playing Y if the signal is nb.

In case c, α and β are commonly known, the extensive form game $\Gamma(\alpha)$ is described in figure 2.



$c > 0,\ 0 < r_i < w_i < 1,\ i = 1, 2$

Figure 2: The Game Γ

It is assumed that the precision α of IS is 2's private information and its distribution is common knowledge. It is given by a continuous p.d.f $f(\alpha) : (\frac{1}{2}, 1) \rightarrow \Re_+$, $\int_{\frac{1}{2}}^{1} f(\alpha) d\alpha = 1$. The cost, c, of allowing inspection is 1's private information and its distribution is commonly known. For NB to be not strictly dominated by O for all types of 1, it is assumed $Pr(c > w_1 - r_1) > 0$. Finally, β is assumed to be common knowledge. That is, Player 2 (the strong nation) is well informed about Player 1's (the weak nation) disruptive technology. Let Γ be the Bayesian game described above.
We restrict our analysis to Perfect Bayesian Equilibrium (PBE) of Γ.

Claim 1. *Given that 1 does not open his facilities,*

1. *In Γ_1, the strategy (NA, A) is strictly dominated by (A, NA) for all types α, $\frac{1}{2} < \alpha < 1$, of Player 2.*

2. *In a PBE of Γ, Player 1 of any type c plays either pure O or he mixes B and NB.*

Claim 1 asserts that in Γ_1, irrespective of 2's type α, acting opposite to IS's recommendation is strictly dominated by the strategy of acting according to IS's recommendation. In Γ, mixing O and B, mixing O and NB or mixing all three pure strategies can not be a PBE outcome.

Proof: See Appendix.

Consider an equilibrium where Player 1 plays B with probability $0 < p < 1$ and plays NB with probability $1-p$. Given (α, β) and the signal of IS, the conditional probabilities 2 assigns to Player 1 choosing B or NB are respectively

$$P_2(B|\alpha,\beta,b) = \frac{\alpha' p}{\alpha' p + (1-\alpha)(1-p)}, \quad P_2(NB|\alpha,\beta,b) = \frac{(1-\alpha)(1-p)}{\alpha' p + (1-\alpha)(1-p)}$$
$$P_2(B|\alpha,\beta,nb) = \frac{(1-\alpha')p}{(1-\alpha')p + \alpha(1-p)}, \quad P_2(NB|\alpha,\beta,nb) = \frac{\alpha(1-p)}{(1-\alpha')p + \alpha(1-p)}$$

where $\alpha' = \frac{\beta}{2} + (1-\beta)\alpha$. Player 2's conditional expected payoffs are

$$\Pi_2(A|\alpha,\beta,b) = \frac{\alpha' p\, w_2 + (1-\alpha)(1-p)\, r_2}{\alpha' p + (1-\alpha)(1-p)} \tag{1}$$

$$\Pi_2(NA|\alpha,\beta,b) = \frac{(1-\alpha)(1-p)}{\alpha' p + (1-\alpha)(1-p)} \tag{2}$$

$$\Pi_2(A|\alpha,\beta,nb) = \frac{(1-\alpha')p\, w_2 + \alpha\,(1-p)\, r_2}{(1-\alpha')p + \alpha(1-p)} \tag{3}$$

$$\Pi_2(NA|\alpha,\beta,nb) = \frac{\alpha(1-p)}{(1-\alpha')p + \alpha(1-p)} \tag{4}$$

Given (α, β, p), by (1) and (2),

$$A|b \succ_2 NA|b \longleftrightarrow \alpha > \lambda(p,\beta) \tag{5}$$

where

$$\lambda(p,\beta) \equiv \frac{(1-p)(1-r_2) - \frac{1}{2}\beta p w_2}{(1-p)(1-r_2) + (1-\beta)p w_2} \tag{6}$$

Namely, Player 2 after observing the signal b, strictly prefers A to NA iff $\alpha > \lambda(p,\beta)$, and he is indifferent between A and NA if $\alpha = \lambda(p,\beta)$.
Similarly, by (3) and (4),

$$A|nb \succ_2 NA|nb \longleftrightarrow \alpha < \frac{(1-\frac{1}{2}\beta)p w_2}{(1-p)(1-r_2) + (1-\beta)p w_2} = 1 - \lambda(p,\beta) \tag{7}$$

Namely, Player 2, after observing the signal nb, strictly prefers A to NA iff $\alpha < 1-\lambda(p,\beta)$ and he is indifferent between A and NA if $\alpha = 1-\lambda(p,\beta)$.
Notice that by (6), $\lambda(p,\beta)$ is decreasing in p and

$$\left.\begin{array}{ll} \lambda(p,\beta) \in (\dfrac{1}{2}, 1) & \text{iff } 0 < p < \dfrac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \\ 1-\lambda(p,\beta) \in (\dfrac{1}{2}, 1) & \text{iff } \dfrac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} < p < \dfrac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \end{array}\right\} \tag{8}$$

Let $s_2(x|\alpha,\beta,p)$ be 2's best reply to signal $x \in \{b, nb\}$ given (α,β,p). By (5), (6) and (8), for $\alpha \neq \lambda(p,\beta)$

$$s_2(b|\alpha,\beta,p) = \begin{cases} NA, & \text{if } \ p < \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \ \text{ and } \ \frac{1}{2} < \alpha < \lambda(p,\beta) \\ A, & \text{if } \ p < \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \ \text{ and } \ \lambda(p,\beta) < \alpha < 1 \\ & \text{or } \ p > \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \end{cases} \tag{9}$$

By (7) and (8), for $\alpha \neq 1-\lambda(p,\beta)$

$$s_2(nb|\alpha,\beta,p) = \begin{cases} NA, & \text{if } \ p < \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \ \text{ and } \ \frac{1}{2} < \alpha < 1 \\ & \text{or } \ \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \leq p < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \ \text{ and } \ 1-\lambda(p,\beta) < \alpha < 1 \\ A, & \text{if } \ \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \leq p < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \ \text{ and } \ \frac{1}{2} < \alpha < 1-\lambda(p,\beta) \\ & \text{or } \ \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \leq p < 1 \ \text{ and } \ \frac{1}{2} < \alpha < 1 \end{cases} \tag{10}$$

Combining (9) and (10), Player 2's best reply strategy as a function of 1's strategy $(p, 1-p)$ and of $(\alpha,\beta,x), x \in \{b, nb\}$ is

$$br_2(p,\alpha,\beta,x) = \begin{cases} NA(\text{ irrespective of signal}), & p \leq \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \text{ and } \alpha < \lambda(p,\beta) \\ A(\text{ irrespective of signal}), & \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} < p < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \text{ and } \alpha < 1-\lambda(p,\beta) \\ & \text{or } p > \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \\ NA \text{ if } x = nb, A \text{ if } x = b & p \leq \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} \text{ and } \alpha > \lambda(p,\beta) \\ & \text{or } \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2} < p < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2} \text{ and } \alpha > 1-\lambda(p,\beta) \end{cases} \tag{11}$$

For a fixed β, $br_2(p,\alpha,\beta,x)$ is illustrated in Figure 3 below.

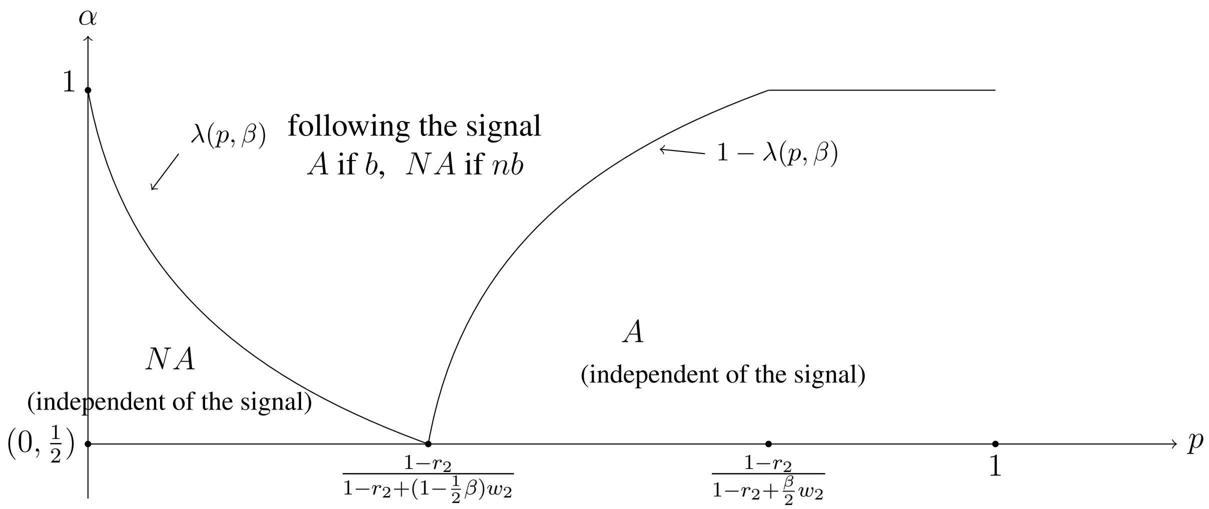
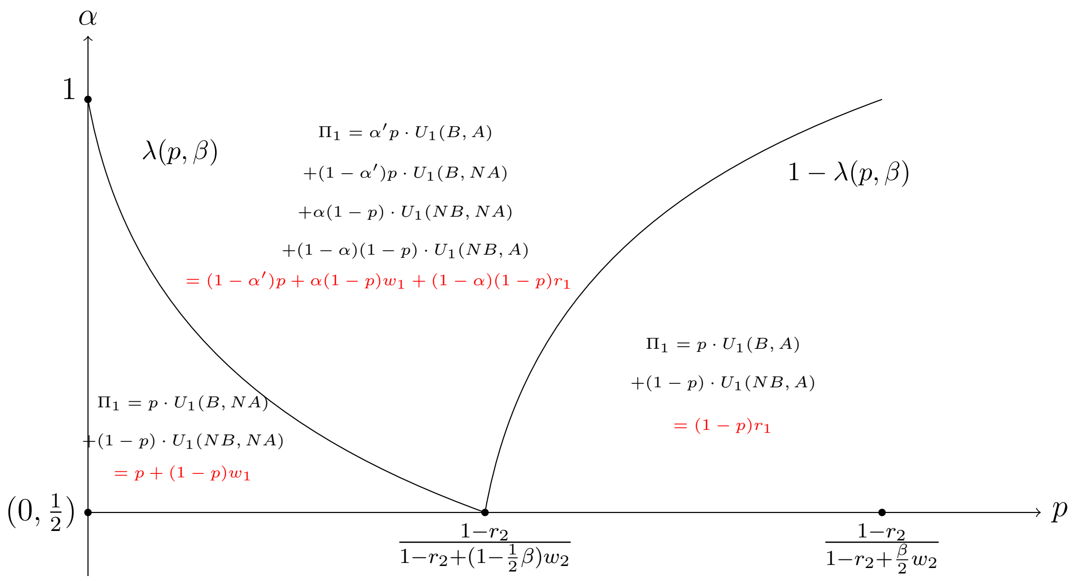

α
1
$\lambda(p,\beta)$
following the signal
A if b, NA if nb
$1-\lambda(p,\beta)$
A
(independent of the signal)
NA
(independent of the signal)
$(0,\frac{1}{2})$
$\frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$
$\frac{1-r_2}{1-r_2+\frac{\beta}{2}w_2}$
1
p

Figure 3: Player 2's best reply $br_2(p|\alpha,\beta,x)$

Claim 2. *In equilibrium* $p^* < \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$.

Proof: Suppose to the contrary, $p^* \geq \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}$. Then every type α of Player 2 attacks 1 irrespective of the signal, thus 1 is better off playing pure NB. But then 2 is better off deviating and not attack, a contradiction. □

The ex-post expected payoff of Player 1 (i.e. his expected payoff as a function of α) given 2's best reply strategy is illustrated in Figure 4 below


α
1
$\lambda(p,\beta)$
$\Pi_1 = \alpha' p \cdot U_1(B,A)$
$+(1-\alpha')p \cdot U_1(B,NA)$
$+\alpha(1-p) \cdot U_1(NB,NA)$
$+(1-\alpha)(1-p) \cdot U_1(NB,A)$
$= (1-\alpha')p + \alpha(1-p)w_1 + (1-\alpha)(1-p)r_1$
$1-\lambda(p,\beta)$
$\Pi_1 = p \cdot U_1(B,A)$
$+(1-p) \cdot U_1(NB,A)$
$= (1-p)r_1$
$\Pi_1 = p \cdot U_1(B,NA)$
$+(1-p) \cdot U_1(NB,NA)$
$= p + (1-p)w_1$
$(0, \frac{1}{2})$
$\frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$
$\frac{1-r_2}{1-r_2+\frac{\beta}{2}w_2}$
p

$$U_1(B,A)=0, \quad U_1(B,NA)=1, \quad U_1(NB,A)=r_1, \quad U_1(NB,NA)=w_1$$

Figure 4: Player 1's ex-post expected payoff as a function of α and p given $br_2(p,\alpha,\beta,x)$

Let

$$\bar{\alpha} \equiv \frac{1-r_1-\frac{1}{2}\beta}{1+w_1-r_1-\beta}$$

be the threshold which distinguishes a "better quality" IS ($\alpha \geq \bar{\alpha}$) from a "lesser quality" IS ($\alpha < \bar{\alpha}$). Note that $\bar{\alpha} > \frac{1}{2}$ iff $r_1 < 1-w_1$, and $\bar{\alpha} < 1$ iff $\frac{\beta}{2} < w_1$.

Definition. *(i) Player 1 of type c acts more aggressively if he plays B with higher probability.*
(ii) For Player 2 of type α, the strategy (A,A) (Attacking irrespective of signal) is more aggressive than (A,NA) (Following the signal). The strategy (A,NA) is more aggressive than (NA,NA) (Not Attacking irrespective of signal).
(iii) Player 2 acts more aggressively with respect to a change of a parameter if the equilibrium strategy of 2 of all type α, $\frac{1}{2} < \alpha < 1$, is either more aggressive or remains unchanged, and for some α, it is more aggressive.

Proposition 1. *The game Γ has a unique Bayesian perfect equilibrium. It satisfies*
(1.) Every type α of player 2 plays a pure strategy.
(2.) There exists $\bar{c} > 0$ s.t. Player 1 of type c, $0 < c < \bar{c}$, plays the pure strategy O. Player 1 of type c, $c > \bar{c}$, mixes the two strategies B and NB.
(3.) Suppose $c > \bar{c}$ and 1 plays $(p^*, 1-p^*)$, $0 < p^* < 1$.

(3.1) Suppose $E(\alpha) \geq \bar{\alpha}$. Then $p^* = \bar{p_1} \in \left(0, \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}\right]$. Following the signal b, Player 2 attacks 1 iff $\alpha > \lambda(\bar{p_1}, \beta)$. Following the signal nb, Player 2 of any type α, $\frac{1}{2} < \alpha < 1$, does not attack 1.

(3.2) Suppose $E(\alpha) < \bar{\alpha}$. Then $p^* = \bar{p_2} \in \left(\frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}, \frac{1-r_2}{1-r_2+\frac{\beta}{2}w_2}\right)$. Following the signal b, Player 2 of any type α, $\frac{1}{2} < \alpha < 1$, attacks 1. Following the signal nb, Player 2 attacks 1 iff $\frac{1}{2} < \alpha < 1 - \lambda(\bar{p_2}, \beta)$.

(3.3) The better is the disruptive technology of 1, the more aggressively Player 2 acts, and the lower is Player 1's expected payoff.

(4.) The better is the disruptive technology, the higher is the probability that 1 opens his facilities for inspection and avoids the potential attack.

Proof: See Appendix.

Proposition 1 characterizes the Perfect Bayesian Equilibrium of Γ. Player1 opens his facilities and 2 does not attack him if his cost of being inspected is relatively low. Otherwise, 1 refuses to open and he mixes the two pure strategies: building a bomb (B) and not building a bomb (NB).

Suppose 1 estimates IS to be of better quality ($E(\alpha) \geq \bar{\alpha}$), He [1] anticipates Player 2 to detect his action with significant probability, and he builds a bomb with a low probability ($\bar{p_1} \leq \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$). Taking this into account, Player 2 of type $\alpha < \lambda(\bar{p_1})$ (a less accurate IS) regards the signal b as unreliable and does not attack to avoid a possible mistake of attacking an innocent Player 1. Therefore, if the signal is b, only Player 2 of type $\alpha > \lambda(\bar{p_1})$ follows the signal and attacks. If the signal is nb, all types of Player 2 do not attack.

Suppose 1 estimates IS to be a lesser quality ($E(\alpha) < \bar{\alpha}$). He[2] anticipates the low quality IS to send the wrong signal with significant probability and hence not to be detected of building a bomb. As a result, he builds a bomb with high probability ($\bar{p_2} > \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$). The best reply strategy of Player 2 of type $\alpha < 1 - \lambda(\bar{p_2})$ (a less accurate IS) is to treat the signal nb as unreliable and ignore it. Thus she attacks 1 to avoid missing a bomber. Only Player 2 of type $\alpha > 1 - \lambda(\bar{p_2})$ follows the signal nb and does not attack. If the signal is b, all types of Player 2 attack.

Interestingly, the smaller is 1's estimate on IS's quality, the more aggressively 2 behaves. Proposition 1 asserts quite surprisingly that better DT lowers Player 1's expected payoff. The higher is β, the more aggressively Player 2 (who knows β) acts. If $E(\alpha) \geq \bar{\alpha}$, higher β increases the probability of 1 playing B, which induces 2 to attack 1 with higher probability (i.e. attacking 1 for even lesser precision of IS), thereby reducing the payoff of 1. If $E(\alpha) < \bar{\alpha}$, 1 builds a bomb with relatively high probability and if IS's precision is below a certain

[1] Note that $E(\alpha) \geq \bar{\alpha}$ can hold only if $\bar{\alpha} < 1$, or equivalently $\frac{\beta}{2} > w_1$. If $\frac{\beta}{2} < w_1$, then $p^* = \bar{p_2} > \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$.

[2] Note that $E(\alpha) < \bar{\alpha}$ can hold only if $\bar{\alpha} > \frac{1}{2}$. If $\bar{\alpha} \leq \frac{1}{2}$, i.e. $r_1 \geq 1 - w_1$, then the equilibrium strategy p^* must be $\bar{p_1} \leq \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}$. The condition $r_1 \geq 1 - w_1$ is equivalent to $U_1(NB, A) - U_1(B, A) > U_1(B, NA) - U_1(NB, NA)$. The left hand side is the cost of "mistakenly" building a bomb given that 2 attacks; the right hand side is the cost of "mistakenly" not building a bomb given that 2 does not attack. Hence the equilibrium strategy of 1 is never $\bar{p_2}$ if the former "mistake" is more costly than the latter.

threshold, 2 will attack 1 irrespective of the signal. A higher β increases this threshold and therefore, for a wider range of α, Player 2 attacks 1 regardless of the signal. In both cases, Player 2 acts more aggressively, reducing Player 1's expected payoff. This suggests 1 is best of scrapping his DT if he can credibly commit to it. Finally, not less surprisingly, the probability of 1 opening his facility for inspection is increasing with the quality of DT.

We next analyze 2's payoff: If 1 chooses O, Player 2's payoff is 1. Taking into account the best reply strategy of 2 (see (11) and Figure 3), by Proposition 1 Player 2's expected payoff in case Player 1 refuses inspection is

$$E\Pi_2^*(\alpha) = \begin{cases} 1-\bar{p_1}, & \text{if } E(\alpha) \geq \bar{\alpha}, \frac{1}{2} < \alpha < \lambda(\bar{p_1}, \beta) \\ \alpha\left[\bar{p_1}w_2(1-\beta) + (1-\bar{p_1})(1-r_2)\right] + (1-\bar{p_1})r_2 + \frac{\beta}{2}\bar{p_1}w_2, & \\ & \text{if } E(\alpha) \geq \bar{\alpha}, \lambda(\bar{p_1}, \beta) \leq \alpha < 1 \\ \alpha\left[\bar{p_2}w_2(1-\beta) + (1-\bar{p_2})(1-r_2)\right] + (1-\bar{p_2})r_2 + \frac{\beta}{2}\bar{p_2}w_2, & \\ & \text{if } E(\alpha) < \bar{\alpha}, 1-\lambda(\bar{p_2}, \beta) \leq \alpha < 1 \\ \bar{p_2}w_2 + (1-\bar{p_2})r_2, & \text{if } E(\alpha) < \bar{\alpha}, \frac{1}{2} < \alpha < 1-\lambda(\bar{p_2}, \beta) \end{cases}$$

Let

$$\hat{\alpha}(p^*, \beta) = \begin{cases} \lambda(\bar{p_1}, \beta) & \text{if } E(\alpha) \geq \bar{\alpha} \\ 1-\lambda(\bar{p_2}, \beta) & \text{if } E(\alpha) < \bar{\alpha} \end{cases}$$

$\hat{\alpha}$ is the threshold which determines whether Player 2 follows the signal (if $\alpha > \hat{\alpha}$) or not (if $\alpha < \hat{\alpha}$).

Proposition 2. *Consider the perfect Bayesian equilibrium of* Γ.

1. *Suppose $c < \bar{c}$. The outcome is (O, NA) and Player 2 obtains 1.*
2. *Suppose $c > \bar{c}$.*

   If $\alpha < \hat{\alpha}$, Player 2 of type α ignores the signal and her payoff is constant in α.

   If $\alpha > \hat{\alpha}$, Player 2 of type α follows the signal and her payoff is strictly increasing in α .

Proposition 2 asserts that Player 2's payoff is non-decreasing in α. It is strictly increasing if $\alpha > \hat{\alpha}$ so that she trust the signal and acts accordingly. If $\alpha < \hat{\alpha}$, Player 2 ignores the signal and her payoff is independent of α. In this case there are two scenarios: (1) $p^* = \bar{p_1}$ and $\alpha < \lambda(\bar{p_1}, \beta)$ or (2) $p^* = \bar{p_2}$ and $\alpha < 1-\lambda(\bar{p_2}, \beta)$. In the former case, with positive probability($\alpha'\bar{p_1}$), Player 1 possesses a bomb, IS sends the correct signal b, but Player 2 misses attacking 1. In the latter case, with positive probability ($(1-\bar{p_2})\alpha$), Player 1 does not build a bomb, IS sends the correct signal nb, but Player 2 mistakenly attacks the innocent Player 1. For a significantly wide range of α, these two tragic events are likely to occur.

Next we study the effect of IS's precision on Player 1's ex-post equilibrium payoff, $\Pi_1^*(\alpha)$.

Proposition 3. *Suppose $c > \bar{c}$.*
In the perfect Bayesian equilibrium, $\Pi_1^*(\alpha)$ is constant in $\alpha \in \Big(\frac{1}{2}, \hat{\alpha}\Big)$. In the interval $\Big(\hat{\alpha}, 1\Big)$, if $p^* < \frac{w_1 - r_1}{1+w_1-r_1-\beta}$, $\Pi_1^*(\alpha)$ is strictly increasing in α, and if $p^* > \frac{w_1-r_1}{1+w_1-r_1-\beta}$, it is strictly decreasing in α.

Proof: See Appendix.

By Proposition 3, if 2 follows the signal and 1 plays B with high probability ($p^* > \frac{w_1-r_1}{1+w_1-r_1-\beta}$), 1 is better off with less accurate IS since it lowers the chance 2 receiving the correct signal b and missing attacking 1. On the other hand, when 2 follows the signal and 1 plays B with low probability ($p^* < \frac{w_1-r_1}{1+w_1-r_1-\beta}$), 1 prefers more accurate IS to increase the likelyhood of receiving the signal nb and not attacking.

Next we analyze the effect of Player 1's estimate of IS's precision ($E(\alpha)$) on his ex-post expected payoff.
Suppose $c > \bar{c}$. Using Figure 4 above and Proposition 1, 1's ex-post expected payoff is

$$\Pi_1^*(\alpha) = \begin{cases} \Pi_{1,1}^* &= (1-w_1)\bar{p_1} + w_1, \qquad \text{if } E(\alpha) \geq \bar{\alpha} \text{ and } \frac{1}{2} < \alpha < \lambda(\bar{p_1}, \beta) \\ \Pi_{1,2}^* &= \Big[1 - \frac{\beta}{2} - r_1 - (1 + w_1 - r_1 - \beta)\alpha\Big]\bar{p_1} + (w_1 - r_1)\alpha + r_1, \\ & \qquad \text{if } E(\alpha) \geq \bar{\alpha} \text{ and } \lambda(\bar{p_1}, \beta) \leq \alpha < 1 \\ \Pi_{1,3}^* &= \Big[1 - \frac{\beta}{2} - r_1 - (1 + w_1 - r_1 - \beta)\alpha\Big]\bar{p_2} + (w_1 - r_1)\alpha + r_1, \\ & \qquad \text{if } E(\alpha) < \bar{\alpha} \text{ and } 1 - \lambda(\bar{p_2}, \beta) \leq \alpha < 1 \\ \Pi_{1,4}^* &= -r_1\bar{p_2} + r_1, \qquad \text{if } E(\alpha) < \bar{\alpha} \text{ and } \frac{1}{2} < \alpha < 1 - \lambda(\bar{p_2}, \beta) \end{cases} \tag{12}$$

It is easy to verify that

$$\Pi_{1,1}^* - \Pi_{1,2}^* = \alpha\bar{p_1}(1-\beta) + \frac{\beta}{2}\bar{p_1} + (1-\bar{p_1})(1-\alpha)(w_1 - r_1) > 0$$

and

$$\Pi_{1,3}^* - \Pi_{1,4}^* = \alpha(w_1 - r_1)(1-\bar{p_2}) + \bar{p_2}[(1-\alpha) + \beta(\alpha - \frac{1}{2})] > 0$$

Case 1. Suppose $\alpha > \bar{\alpha}$.

If 1's estimate is consistent with the actual precision (namely, $E(\alpha) \geq \bar{\alpha}$), his payoff, by (12), is either $\Pi_{1,1}^*$ or $\Pi_{1,2}^*$. If, however, 1 underestimates IS to be of lesser quality ($E(\alpha) < \bar{\alpha}$), his payoff is either $\Pi_{1,3}^*$ or $\Pi_{1,4}^*$. Since $\alpha > \bar{\alpha}$ is equivalent to $1 - \frac{\beta}{2} - r_1 - (1 + w_1 - r_1 - \beta)\alpha < 0$, and since $\bar{p_1} < \bar{p_2}$ (see Proposition 1), we have, by (12), $\Pi_{1,2}^* > \Pi_{1,3}^*$. Thus both $\Pi_{1,1}^*$ and $\Pi_{1,2}^*$ are greater than the maximum of $\Pi_{1,3}^*$ and $\Pi_{1,4}^*$. Consequently, 1's payoff is higher if his estimate is consistent with the actual precision of IS.

Case 2. Suppose $\alpha < \bar{\alpha}$.

If 1's estimate is consistent with the actual precision (namely, $E(\alpha) < \bar{\alpha}$), his payoff is either $\Pi^*_{1,3}$ or $\Pi^*_{1,4}$. If 1 overestimates IS ($E(\alpha) \geq \bar{\alpha}$), his payoff is either $\Pi^*_{1,1}$ or $\Pi^*_{1,2}$. Since $\alpha < \bar{\alpha}$ is equivalent to $1 - \frac{\beta}{2} - r_1 - (1 + w_1 - r_1 - \beta)\alpha > 0$, by (12), $\Pi^*_{1,2} < \Pi^*_{1,3}$, and a consistent estimate benefits 1. On the other hand, since

$$\Pi^*_{1,1} - \Pi^*_{1,4} = (w_1 - r_1) + (1 - w_1)\bar{p_1} + r_1\bar{p_2} > 0$$

The conclusion regarding whether or not 1 benefits from a consistent estimation of IS is ambiguous.

We summarize the above in the next proposition.

Proposition 4. *Suppose $c > \bar{c}$ and $\alpha > \bar{\alpha}$. Then $\Pi^*_1(\alpha|E(\alpha) \geq \bar{\alpha}) > \Pi^*_1(\alpha|E(\alpha) < \bar{\alpha})$. Suppose $c > \bar{c}$ and $\alpha < \bar{\alpha}$. Then, depending on parameters, $\Pi^*_1(\alpha|E(\alpha) \geq \bar{\alpha}) - \Pi^*_1(\alpha|E(\alpha) < \bar{\alpha})$ can be either positive or negative.*

4 Example: Uniform Distribution

In this section, we illustrate the results for the case where α is uniformly distributed. That is, $f(\alpha) = 2$ for $\alpha \in (\frac{1}{2}, 1)$. In equilibrium, if $c < \bar{c}$, Player 1 opens facilities for inspection and Player 2 does not attack 1. If $c > \bar{c}$, Player 1 builds a bomb with probability p^*, where

$$p^* = \frac{\left[2 - \beta - \sqrt{(\beta - 2w_1)^2 + 4(1 - w_1)r_1}\right](1 - r_2)}{\left[(1 - \beta)w_2 + r_2 - 1\right]\sqrt{(\beta - 2w_1)^2 + 4(1 - w_1)r_1} + \left[(w_1 - r_1)w_2 + 1 - r_2\right](2 - \beta)}$$

$$\bar{c} = \frac{(w_1 - r_1)\left[2 - \beta - \sqrt{(\beta - 2w_1)^2 + 4(1 - w_1)r_1}\right]^2}{4\,(1 - \beta + w_1 - r_1)^2}$$

1. If $3w_1 - (1 - r_1) > \beta$, (namely, $E(\alpha) > \bar{\alpha}$) and $c > \bar{c}$. Following the signal nb, Player 2 does not attack 1 irrespective of α. Following the signal b, 2 still does not attack 1 if

$$\alpha < \hat{\alpha_1} = \frac{2(w_1 - r_1) - \beta + \sqrt{(\beta - 2w_1)^2 + 4(1 - w_1)r_1}}{2\,(1 - \beta + w_1 - r_1)}$$

   Only if $\alpha > \hat{\alpha_1}$, 2 relies on IS and attacks 1 when observing b.

2. If $3w_1 - (1 - r_1) < \beta$, (namely, $E(\alpha) < \bar{\alpha}$) and $c > \bar{c}$. Following the signal b, Player 2 attacks 1 irrespective of α. Following the signal nb, 2 still attacks 1 if

$$\alpha < \hat{\alpha_2} = \frac{2 - \beta - \sqrt{(\beta - 2w_1)^2 + 4(1 - w_1)r_1}}{2\,(1 - \beta + w_1 - r_1)}$$

   Only if $\alpha > \hat{\alpha_2}$, 2 relies on IS and does not attack 1 when observing nb.

	Example 1.	Example 2.	Example 3.	Example 4.
	$\beta = 0$, $r_1 = 0.1, w_1 = 0.3$, $r_2 = 0.2, w_2 = 0.4$	$\beta = 0.5$, $r_1 = 0.1, w_1 = 0.3$, $r_2 = 0.2, w_2 = 0.4$	$\beta = 0.5$, $r_1 = 0.1, w_1 = 0.7$, $r_2 = 0.2, w_2 = 0.4$	$\beta = 0.9$, $r_1 = 0.1, w_1 = 0.7$, $r_2 = 0.2, w_2 = 0.4$
$\bar{c}$	0.05	0.094	0.036	0.074
p^*	0.667	0.772	0.437	0.577
$\hat{\alpha}$	0.5	0.687	0.757	0.649
$E\Pi_1^*$	0.25	0.206	0.664	0.626
$\Pi_1(\alpha)$ if $\alpha < \hat{\alpha}$		0.023	0.831	0.873
$\Pi_1(\alpha)$ if $\alpha > \hat{\alpha}$	$-0.6\alpha + 0.7$	-0.34α+0.6	0.12α+0.384	0.196α+0.36

In Example 1, Player 1 estimates the precision of IS to be $\bar{\alpha}$. Player 2 of all types follows the signal, and 1's ex post payoff is decreasing in α.
Increasing β only, we move to Example 2, where 1 plays B with higher probability and lowers his expected payoff from 0.25 to 0.206. If $\alpha < 0.687$, Player 2 attacks. If $\alpha > 0.687$, Player 2 follows the signal. 1's ex post payoff is non-increasing in α.
Increasing w_1 only, we move to Example 3, where p^* changes from 0.772 ($p^* = \bar{p_2}$) to 0.437 ($p^* = \bar{p_1}$). In this example, if $\alpha < 0.757$, Player 2 does not attack; if $\alpha > 0.757$, Player 2 follows the signal. 1's ex post payoff is non-decreasing in α.
Increasing β only, we move to Example 4. 1's expected payoff again decreases from 0.664 to 0.626.

5 Appendix

Proof of Claim 1

Proof. (1.) Suppose 1 plays NB in Γ_1. Player 2 of type α obtains $1 - \alpha' + \alpha r_2$ by playing (NA, A), and $(1 - \alpha)r_2 + \alpha$ by playing (A, NA). Since $\alpha > \frac{1}{2}$, 2 prefers to play (A, NA).
Suppose 1 plays B in Γ_1. Player 2 of type α obtains $(1 - \alpha')w_2$ by playing (NA, A), and $\alpha' w_2$ by playing (A, NA). Since $\alpha' = \frac{\beta}{2} + (1 - \beta)\alpha > \frac{1}{2}$, 2 prefers to play (A, NA).
(2.) Suppose, to the contrary, 1 in equilibrium plays either pure B or mixes O and B. If Player 1 refuses to open, Player 2 knows 1 possesses a bomb. The best reply of each type of 2 is to Attack 1. But in this case, Player 1 is better off deviating to NB, a contradiction.
Suppose 1 in equilibrium plays pure NB or mixes O and NB. The best reply of each type of 2 is not to attack 1. In this case, Player 1 is better off deviating to B, a contradiction.

□

Proof of Proposition 1.

In equilibrium 1 plays B with probability $p \in \left(0, \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}\right)$. We distinguish two cases.

Case 1.

Suppose first in equilibrium 1 plays B with probability $p \in \Big(0, \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}\Big]$. Taking into account Player 2's best reply, 1's expected payoff is

$$\begin{aligned}&E_\alpha \Pi_1(p, br_2(p))\\ =&\int_{\frac{1}{2}}^{\lambda(p,\beta)} \Big[p+(1-p)\,w_1\Big] f(\alpha)d\alpha + \int_{\lambda(p,\beta)}^{1} \Big\{(1-\alpha')p+(1-p)\;[\alpha w_1+(1-\alpha)r_1]\;\Big\} f(\alpha)d\alpha\end{aligned} \tag{13}$$

Note that Player 2's action depends only on the signal sent by the IS and its precision. Hence Player 2 is unable to distinguish a deviation of Player 1 from his mixed strategy $(p, 1-p)$ to any other mixture of B and NB (including pure B and pure NB), so 2's strategy remains the same. In equilibrium, since $p \in (0, 1)$, 1 is indifferent with playing $(p, 1-p)$ or playing either one of his pure strategies. In particular, if Player 1 unilaterally deviates to pure B and Player 2 of any type sticks to her best reply strategy against $(p, 1-p)$, 1's expected payoff is

$$E_\alpha \Pi_1(1, br_2(p)) = \int_{\frac{1}{2}}^{\lambda(p,\beta)} f(\alpha)d\alpha + \int_{\lambda(p,\beta)}^{1} (1-\alpha')f(\alpha)d\alpha \tag{14}$$

By equating $E_\alpha \Pi_1(p, br_2(p))$ and $E_\alpha \Pi_1(1, br_2(p))$, we have

$$\int_{\frac{1}{2}}^{\lambda(p,\beta)} (1-w_1)f(\alpha)d\alpha = \int_{\lambda(p,\beta)}^{1} \Big[(\frac{1}{2}\beta+r_1-1)+\alpha(1-\beta+w_1-r_1)\Big] f(\alpha)d\alpha$$

Define for $x \in [\frac{1}{2}, 1]$,

$$m(x) \equiv \int_{\frac{1}{2}}^{x} (1-w_1)f(\alpha)d\alpha - \int_{x}^{1} \Big[(\frac{1}{2}\beta+r_1-1)+\alpha(1-\beta+w_1-r_1)\Big] f(\alpha)d\alpha \tag{15}$$

Clearly, $m_1(x)$ is continuous and differentiable.

$$\begin{aligned}m(\frac{1}{2}) &= -\Big[\frac{1}{2}\beta+r_1-1+(1-\beta+w_1-r_1)E(\alpha)\Big]\\ m(1) &= 1-w_1 > 0\\ m'(x) &= (1+w_1)f(x)-(\frac{1}{2}\beta+r_1-1)f(x)-(1-\beta+w_1-r_1)xf(x)\end{aligned}$$

and

$$m'(x) > 0 \text{ iff } (2-w_1-r_1-\frac{1}{2}\beta)-(1+w_1-r_1-\beta)x > 0$$

which always holds. If $E(\alpha) \geq \frac{1-r_1-\frac{1}{2}\beta}{1+w_1-r_1-\beta}$, then $m(\frac{1}{2}) \leq 0$, $m(1) > 0$, and $m(x)$ is increasing. Hence $m(x) = 0$ has a unique solution $x_1 \in [\frac{1}{2}, 1)$. Since $\lambda(p) \in [\frac{1}{2}, 1)$ is monotonic, there exists $\bar{p_1}$ s.t. $\lambda(\bar{p_1}) = x_1$. Then $\bar{p_1}$ is in $\Big(0, \frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}\Big]$ s.t. $m(\lambda(\bar{p_1}, \beta)) = 0$.

$$\frac{\partial m}{\partial \beta}(\lambda(\bar{p_1}, \beta)) = 0 \longleftrightarrow$$

$$\frac{\partial \lambda(\bar{p_1}, \beta)}{\partial \beta} \cdot f(\lambda(\bar{p_1}, \beta)) \cdot \Big[\lambda(\bar{p_1}, \beta)\,(1-\beta+w_1-r_1) + \frac{\beta}{2} + r_1 - w_1\Big] = -\int_{\lambda(\bar{p_1},\beta)}^{1} (\alpha - \frac{1}{2}) f(\alpha) d\alpha$$

Since $\lambda(\bar{p_1}, \beta)\,(1-\beta+w_1-r_1) + \frac{\beta}{2} + r_1 - w_1 > \frac{1}{2}\,(1-\beta+w_1-r_1) + \frac{\beta}{2} + r_1 - w_1 > 0$, we have

$$\frac{\partial \lambda(\bar{p_1}, \beta)}{\partial \beta} < 0 \tag{16}$$

Case 2.

Suppose 1 plays in equilibrium (p,$1-p$), $p \in \Big(\frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}, \frac{1-r_2}{1-r_2+\frac{1}{2}\beta w_2}\Big)$. Taking into account the best reply strategy of Player 2 of type α, 1's expected payoff is

$$\begin{aligned} & E_\alpha \Pi_1(p, br_2(p)) \\ = & \int_{\frac{1}{2}}^{1-\lambda(p,\beta)} \Big[(1-p)\,r_1\Big] f(\alpha) d\alpha + \int_{1-\lambda(p,\beta)}^{1} \Big\{(1-\alpha')p + (1-p)\ [\alpha w_1 + (1-\alpha) r_1]\Big\} f(\alpha) d\alpha \end{aligned} \tag{17}$$

If Player 1 deviates to B, and Player 2 sticks to her best reply strategy to $(p, 1-p)$, then 1's expected payoff is

$$E_\alpha \Pi_1(1, br_2(p)) = \int_{1-\lambda(p,\beta)}^{1} (1-\alpha') f(\alpha) d\alpha \tag{18}$$

By equating $E_\alpha \Pi_1(p, br_2(p)) = E_\alpha \Pi_1(1, br_2(p))$, we have

$$\int_{1-\lambda(p,\beta)}^{1} [1 - \frac{1}{2}\beta - (1-\beta)\alpha] f(\alpha) d\alpha = r_1 + (w_1 - r_1) \int_{1-\lambda(p,\beta)}^{1} \alpha f(\alpha) d\alpha$$

Define for $x \in [\frac{1}{2}, 1]$

$$g(x) \equiv \int_{1}^{x} [1 - \frac{1}{2}\beta - (1-\beta+w_1-r_1)\alpha] f(\alpha) d\alpha + r_1 \tag{19}$$

Clearly, $g(x)$ is continuous and differentiable and we have

$$\begin{aligned} g(\frac{1}{2}) &= (1-\beta+w_1-r_1)E(\alpha)-(1-\frac{1}{2}\beta-r_1) \\ g(1) &= r_1 > 0 \\ g'(x) &= [1-\frac{1}{2}\beta-(1-\beta+w_1-r_1)x]f(x) \end{aligned}$$

and $g'(x) > 0$ iff $x < \frac{1-\frac{1}{2}\beta}{1-\beta+w_1-r_1}$.
If $E(\alpha) < \frac{1-r_1-\frac{1}{2}\beta}{1+w_1-r_1-\beta}$, then $g(\frac{1}{2}) < 0$, $g(1) > 0$, and $g(x)$ intersects the x-axis only once in the interval $(\frac{1}{2}, \frac{1-\frac{1}{2}\beta}{1-\beta+w_1-r_1})$. Hence $g(x) = 0$ has a unique solution $x_2 \in (\frac{1}{2}, \frac{1-\frac{1}{2}\beta}{1-\beta+w_1-r_1})$. Since $1-\lambda(p,\beta)$ is monotonic, there exists $\bar{p_2}$ s.t. $1-\lambda(\bar{p_2},\beta) = x_2$. Then $\bar{p_2}$ is in $\left(\frac{1-r_2}{1-r_2+(1-\frac{1}{2}\beta)w_2}, \frac{1-r_2}{w_2(w_1-r_1)+(1-r_2)}\right)$ and is the solution to $g(1-\lambda(\bar{p_2},\beta)) = 0$.

Similar to the previous case,

$$\frac{\partial g}{\partial \beta}(1-\lambda(\bar{p_2})) = 0 \longleftrightarrow$$

$$\frac{\partial(1-\lambda)}{\partial\beta}(\bar{p_2},\beta)\cdot f(1-\lambda(\bar{p_2},\beta))\cdot\Big[1-\frac{\beta}{2}-(1-\lambda(\bar{p_2},\beta))\,(1-\beta+w_1-r_1)\Big] = \int_{1-\lambda(\bar{p_2},\beta)}^{1}(\alpha-\frac{1}{2})f(\alpha)d\alpha$$

Since $1-\frac{\beta}{2}-(1-\lambda(\bar{p_2},\beta))\,(1-\beta+w_1-r_1) > 0$, we have

$$\frac{\partial(1-\lambda)}{\partial\beta}(\bar{p_2},\beta) > 0 \tag{20}$$

If 1 chooses O, his payoff is $w_1 - c$. If 1 refuses to open, Player 1's expected payoff is

$$E_\alpha\Pi_1^* = E_\alpha\Pi_1(0, br_2(p^*)) = \begin{cases} w_1 - \int_{\lambda(\bar{p_1},\beta)}^{1}(1-\alpha)(w_1-r_1)f(\alpha)d\alpha, & \text{if } E(\alpha) \geq \bar{\alpha} \\ r_1 + \int_{1-\lambda(\bar{p_2},\beta)}^{1}\alpha(w_1-r_1)f(\alpha)d\alpha, & \text{if } E(\alpha) < \bar{\alpha} \end{cases} \tag{21}$$

By (16) and (20)

$$\begin{aligned} \frac{\partial E_\alpha\Pi_1^*}{\partial\beta}(\bar{p_1},\beta) &= \Big[1-\lambda(\bar{p_1},\beta)\Big](w_1-r_1)f(\lambda(\bar{p_1},\beta))\cdot\frac{\partial\lambda}{\partial\beta}(\bar{p_1},\beta) < 0 \\ \frac{\partial E_\alpha\Pi_1^*}{\partial\beta}(\bar{p_2},\beta) &= -\Big[1-\lambda(\bar{p_2},\beta)\Big](w_1-r_1)f(1-\lambda(\bar{p_2},\beta))\cdot\frac{\partial(1-\lambda)}{\partial\beta}(\bar{p_2},\beta) < 0 \end{aligned}$$

Therefore, $\dfrac{\partial E_\alpha\Pi_1^*}{\partial\beta}(p^*,\beta) < 0$. Player 1's expected payoff is decreasing in β.

Let

$$\bar{c} = w_1 - E_\alpha \Pi_1^* = \begin{cases} \int_{\lambda(\bar{p_1},\beta)}^1 (1-\alpha)(w_1 - r_1) f(\alpha) d\alpha & \text{if } E(\alpha) \geq \bar{\alpha} \\ w_1 - r_1 - \int_{1-\lambda(\bar{p_2},\beta)}^1 \alpha(w_1 - r_1) f(\alpha) d\alpha & \text{if } E(\alpha) < \bar{\alpha} \end{cases}$$

Player 1 plays pure O if $c < \bar{c}$. He mixes B and NB if $c > \bar{c}$. It is easy to see that $\frac{\partial \bar{c}}{\partial \beta}(p^*, \beta) > 0$ since $\frac{\partial E_\alpha \Pi_1^*}{\partial \beta}(p^*, \beta) < 0$. Therefore $P(c < \bar{c})$ is increasing in β– a better quality of DT raises the probability of 1 opening facilities for inspection. This completes the analysis of the equilibrium strategy of 1. □

Proof of Proposition 3

In the equilibrium where 1 mixes B and NB, his ex-post expected payoff is

$$\Pi_1^*(\alpha) = \begin{cases} (1-\bar{p_1})w_1 + \bar{p_1}, & \text{if } E(\alpha) \geq \bar{\alpha}, \frac{1}{2} < \alpha < \lambda(\bar{p_1}, \beta) \\ [(w_1 - r_1) - (1 + w_1 - r_1 - \beta)\bar{p_1}]\,\alpha + (1-\bar{p_1})r_1 + (1 - \frac{\beta}{2})\bar{p_1}, & \\ & \text{if } E(\alpha) \geq \bar{\alpha}, \lambda(\bar{p_1}, \beta) \leq \alpha < 1 \\ [(w_1 - r_1) - (1 + w_1 - r_1 - \beta)\bar{p_2}]\,\alpha + (1-\bar{p_2})r_1 + (1 - \frac{\beta}{2})\bar{p_2}, & \\ & \text{if } E(\alpha) < \bar{\alpha}, 1 - \lambda(\bar{p_2}, \beta) \leq \alpha < 1 \\ (1-\bar{p_2})r_1, & \text{if } E(\alpha) < \bar{\alpha}, \frac{1}{2} < \alpha < 1 - \lambda(\bar{p_2}, \beta) \end{cases}$$

Suppose $\alpha \in \left(\frac{1}{2}, \hat{\alpha}(p^*, \beta)\right)$. That is, $E(\alpha) \geq \bar{\alpha}$ and $\frac{1}{2} < \alpha < \lambda(\bar{p_1}, \beta)$, or $E(\alpha) < \bar{\alpha}$ and $\frac{1}{2} < \alpha < 1 - \lambda(\bar{p_2}, \beta)$, $\Pi_1^*(\alpha)$ is constant in α.
Suppose $\alpha \in \left(\hat{\alpha}(p^*, \beta), 1\right)$.
Given $E(\alpha) \geq \bar{\alpha}$ and $\lambda(\bar{p_1}, \beta) \leq \alpha < 1$,then $\Pi_1^*(\alpha)$ is increasing in α if $\bar{p_1} < \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$ and decreasing in α if $\bar{p_1} > \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$.
By (6), $\bar{p_1} < \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$ iff $\lambda(\frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}) < \lambda(\bar{p_1}, \beta)$, that is, $\frac{1 - r_2}{(w_1 - r_1)w_2 + 1 - r_2} < \lambda(\bar{p_1}, \beta)$.
Given $E(\alpha) < \bar{\alpha}$ and $1 - \lambda(\bar{p_2}, \beta) \leq \alpha < 1$, then $\Pi_1^*(\alpha)$ is increasing in α if $\bar{p_2} < \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$ and is decreasing in α if $\bar{p_2} > \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$.
By (6), $\bar{p_2} < \frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}$ iff $\lambda(\frac{w_1 - r_1}{1 + w_1 - r_1 - \beta}) < \lambda(\bar{p_2}, \beta)$, that is, $\frac{1 - r_2}{(w_1 - r_1)w_2 + 1 - r_2} < \lambda(\bar{p_2}, \beta)$.
This completes the proof of Proposition 3.

□

Proof of Proposition 4

6 Reference

1. Avenhaus, Rudolf, Bernhard Von Stengel, and Shmuel Zamir. ”Inspection games.” Handbook of game theory with economic applications 3 (2002): 1947-1987.

2. Baliga Sandeep, and Tomas Sjöström. ”Strategic ambiguity and arms proliferation.” Journal of political Economy 116.6 (2008): 1023-1057.

3. Debs Alexandre, and Nuno P. Monteiro. ”Known unknowns: Power shifts, uncertainty, and war.” International Organization 68.1 (2014): 1-31.

4. Jelnov Artyom, Yair Tauman, and Richard Zeckhauser. ”Confronting an enemy with unknown preferences: Deterrer or provocateur?.” European Journal of Political Economy 54 (2018): 124-143.

5. Jelnov Artyom, Yair Tauman, and Richard Zeckhauser. ”Attacking the unknown weapons of a potential bomb builder: The impact of intelligence on the strategic interaction.” Games and Economic Behavior 104 (2017): 177-189.

6. O’Neill, Barry. ”Game theory models of peace and war.” Handbook of game theory with economic applications 2 (1994): 995-1053.