Proportional use of force in counter-terrorism

Artyom Jelnov*

October 18, 2015

Abstract

This paper studies a concept of proportional use of force in a counter-terrorism action. The terrorist organization decides whether to attack or not to attack the state, while the state decides about his counter-terrorist measure. The state is supposed not to use a "non-proportional" force, namely, the counter-terrorist measure taken by the state should be sufficient to remove a threat imposed by terrorists, but not higher than that. The level of force required against the terrorist organization is a private information of the organization. The model predicts under which conditions terrorists will not attack the state, and under which conditions with a positive probability the attack will take place, the state will react with a tough counter-terrorist measure and the state may be accused for non-proportional use of violence.

1 Introduction

In the modern world many states encounter a threat of terrorist attack. I suggest here a model of conflict between a state and a terrorist organization. The state, once attacked by terrorists, has to decide about its counter-measure. It is assumed that the state is military stronger, but its action is limited by moral considerations, international law etc. In particular, state's

*Ariel University, Israel, artyomj@ariel.ac.il

reaction should meet "necessity and proportionality" criteria. This paper is dedicated to the concept of proportional use of force. A question of what is a proportional use of force is widely discussed in the international law literature. For this analysis I adopt the following approach, as stated in Duffy (2015, p. 267): "...Proportionality requiring that the force used be no more than necessary to repel the threat presented". However, in most cases there is no clearness which force is necessary to repeal the threat presented by terrorist organization. In my model the terrorist organization can be one of two types: the "weak" and the "strong". If the terrorist organization is "strong", only large military effort by the state can eliminate its threat. But to repel the threat of the "weak" organization, even low effort by the state is enough. Thus, to use the large force against the "weak" organization might be considered as an excessive, or a non-proportional reaction. I assume that the type of organization is its private information, so the state does not know which threat it meats. Moreover, it should be taken into account that the weak organization may may be interested to provoke the state to use excessive force. The terrorist organization may benefit if the state was accused in a non-proportional reaction. As wrote Wilkinson (2011, p.7), one of goals of terrorism may be "provoking government security forces into over-reaction".

I start with a two-actors model. There are two players: the terrorist organization and the state and its government (since it is assumed that the state's government is a decision maker, both terms "government" and "'state" are synonymous in this paper). The terrorist organization may be "weak" or "strong", and its exogenously given type is its private information. It may attack or not attack the state/government. If there was no attack, the government is not allowed to use the force in this model.[1] It is not the worst outcome for the terrorist organization, for instance since it can threaten the state in the future.

If the government was attacked, it should take some counter-terrorist action. The action can be "weak" or "strong". The "weak" action is sufficient to repel the threat presented

[1] A legitimation of use of preventive force, when a state was not attacked, is a controversy in the international law literature. Many scholars argue that such an action by the state is not legitimate. See Duffy (2015) for a discussion.

by the weak terrorist organization, but is not sufficient against the strong organization. Therefore, if the weak action was taken against the weak organization, it is the best outcome for the government and the worst for the organization. If the weak action was taken against the strong organization, this outcome is better for terrorist organization rather than if it does not attack. It enjoys from benefit of successful attack on the rival government and survived its countermeasure.

The government may also decide to take the strong action as a response to the attack. Then the threat from the terrorist organization of both types will be repelled. But once performed the strong action against the weak organization, the government will be ex-post accused in an over-reaction. I assume here that political consequences of this situation for the government are extremely tough and this is the worst ex-post outcome for the government. Since the terrorist organization's goal is to cause a maximal damage to the government, it is the best outcome for it. Actually, the reason for the weak organization to attack the state is to provoke its government to over-react. On the other hand, if the government took the strong action against the strong organization, this action is considered to be legitimate, and it is successful in the meaning that it eliminates the threat on the government, thus it is the best outcome for the government and the worst for the terrorist organization.

I show that if the payoff of the terrorist organization when it does not attack is sufficiently high, in any Nash equilibrium of this game the organization of both types will not perform an attack with certainty. If the payoff of the organization for the not attacking is not sufficiently high, then the Nash equilibrium is unique, and in the equilibrium the organization attacks with a positive probability. If the government assigns a high probability that the organization is weak, then the strong organization attacks with probability one. And vice verse, if the prior belief that the organization is weak is relatively low, the weak organization attacks for sure. The intuition is, that if the government expects to meet the weak organization with relatively high probability, its reaction on the attack will be weak with higher probability. The strong organization takes advantage on it and attacks. If the government assigns a high probability to being the organization strong, its reaction on the attack will be strong with higher probability, and since the strong action against the weak organization is the best outcome for terrorists, the weak organization will prefer to attack.

This model is described in Section 2.

In the two-actors model the government constraints itself in the sense that it would prefer not to take the strong anti-terrorist action if it knows that the weak action is sufficient. In the next, three-actors model, this self-constraint is removed. The government goal, once attacked, is just to destroy the terrorist organization, even by using an exaggerated force. But there is some third party, for example, the international community, or a domestic juridical system, which checks that the attacked government does not use a non-proportional force. I call this party a Judge.

As in the two-actors model, there are terrorist organization of two possible types, which can attack or not attack, and the government, which reacts weakly or strongly once attacked. If there was no attack, or if the reaction on attack was weak, payoffs of the organization and of the government are as in the previous model.[2] But if the government's reaction was strong, the Judge decides whether to intervene and to punish the government for (probably) over-reaction, or not. If the Judge intervenes and punishes the government, it is the worst outcome for the government and the best for the terrorist organization. If the Judge does not intervene, while the organization is destroyed by the strong anti-terrorist action, and it is the worst outcome for it and the best for the government. As for the Judge, his/her ex-post preferences depend on whether the "justice" was done. Namely, s/he most prefers to intervene if the strong action was taken against the weak organization, and not to intervene in the opposite case.

As in the two-actors game, if the payoff of the terrorist organization when it does not attack is sufficiently high, in any Nash equilibrium of this game the organization of both

[2]For simplicity of analysis I assume here that in the case of no attack, the payoff of the terrorist organization of both types is equal.

types will not perform an attack with certainty. As for the case where the organization's payoff for not attacking is relatively low, multiplicity of Nash equilibria cause to ambiguity of results. To resolve it, I suggest a concept of an "equilibrium compatible with symmetric mistakes". This concept is in spirit of the " trembling-hand perfectness" concept due to Selten (1975)), but has stronger requirements that "trembles of hand" are symmetric for all players and types.

Assume that there is some minimal probability, **symmetric** for all players and types of players, with which each strategy is chosen. It reflects an idea that each strategy, even if it is inferior for some player, may be chosen "by mistake" with some probability, and this probability is symmetric. Then one can find an equilibrium of the new game, when the "symmetric probability of mistakes" is taken into account. If this equilibrium converges to some Nash equilibrium of the initial game, as the probability of mistakes converges to zero, the limit equilibrium is called *compatible with symmetric mistakes.* I argue that in the case of multiplicity of Nash equilibria, a compatible with symmetric mistakes equilibrium, if exists, is a more reasonable one.

The prediction of the model for the case where terrorist organization's payoff when it does not attack is relatively low is that if a prior belief that the terrorist organization is strong is high, then the government will take the strong counter-terrorist measure with a high probability, and the Judge will not intervene with a high probability; taking this into account, the terrorist organization will not attack with certainty. However, if there is high prior belief that the terrorist organization is of the weak type, the government will choose the weak action with a high probability (but less than 1); once it chooses the strong action, the Judge will intervene with a high probability; the terrorist organization of both types will attack with a positive probability. Therefore, it may happen with a positive probability that the strong organization attacks, the government takes the proportional strong counter-measure, but is unjustifiably punished by the third party.

This model appears in Section 3.

The game-theoretic literature on military issues is very rich (see O'Neill (1994) for survey). For a survey of economic approach in studies of terrorism see Enders and Sandler (2011). Lapan and Sandler (1993) study a signalling game model with incomplete information about terrorist resources, but their motivation and model differ from mine. In other researches an incomplete information about terrorists purposes is assumed. For example, see Arce and Sandler (2007).

2 Two-sides game

Consider two players: the first player is a terrorist organization, denoted by T; the second player is a government of the country, threatened by T, and it is denoted by G.

T can be one of two types: the "weak" (W) or the "strong" (S). G assigns a probability p that T is of the type S, and probability $1-p$ that it is W. There are two strategies available for T: to attack the country of G (a) or not to attack it (na) (for simplicity I write hereafter "T attacks/do not attack G"). If he chooses not to attack, the game stops. In this case the payoff of S is β and of W is γ. The payoff of G is λ if it meets W and μ if it meets S.

If T chooses to attack, G has to decide about its reaction counter-terrorist measure. G chooses one of two strategies: the "weak" reaction (w) or the "strong" reaction (s). If T is of the W-type, it is destroyed even by the "weak" reaction, and that is the worst outcome for W and the best outcome for G (it succeed to remove the threat from the terrorist organization). On the other hand, if T is of the S-type, the "weak" counter-terrorist measure is not sufficient to destroy it. This outcome is better for S than if it does not attack: it performed an attack on his enemy country and succeed to survive its reaction. In the latter case $\alpha > \beta$ is the payoff of S and ν is the payoff of G.

If G chooses the "strong" reaction, it destroy the terrorist organization of both types. If the terrorist organization is S, G receives the highest payoff, and S receives the lowest. But if G takes the"strong" measure against the "weak" type of T, it will be accused for

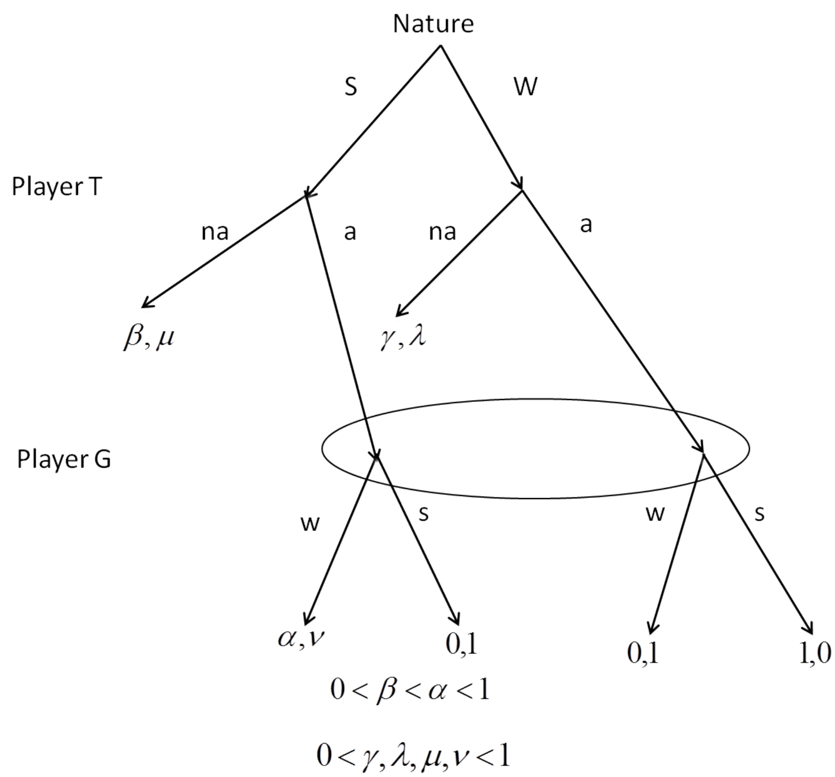


Figure 2.1: The two- players game Γ_2

"overreaction" and "non proportional use of force", and this the worst outcome for G and the best for W.

To summarize, the game, denoted by Γ_2, is presented in the Figure 2.1.

Proposition 2.1. In the game Γ_2, in a Nash equilibrium G mixes w and s strategies and :

(i) if $\beta > \alpha(1-\gamma)$ then in all Nash equilibria T of both types will not attack G with certainty.

(ii) if $\beta < \alpha(1-\gamma)$ and $p < \frac{1}{2-\nu}$ then S attacks G with certainty, while W mixes a and na strategies.

(iii) if $\beta < \alpha(1-\gamma)$ and $p > \frac{1}{2-\nu}$ then W attacks G with certainty, while S mixes a and na strategies.

The proof appears in Appendix.

3 Three-sides game

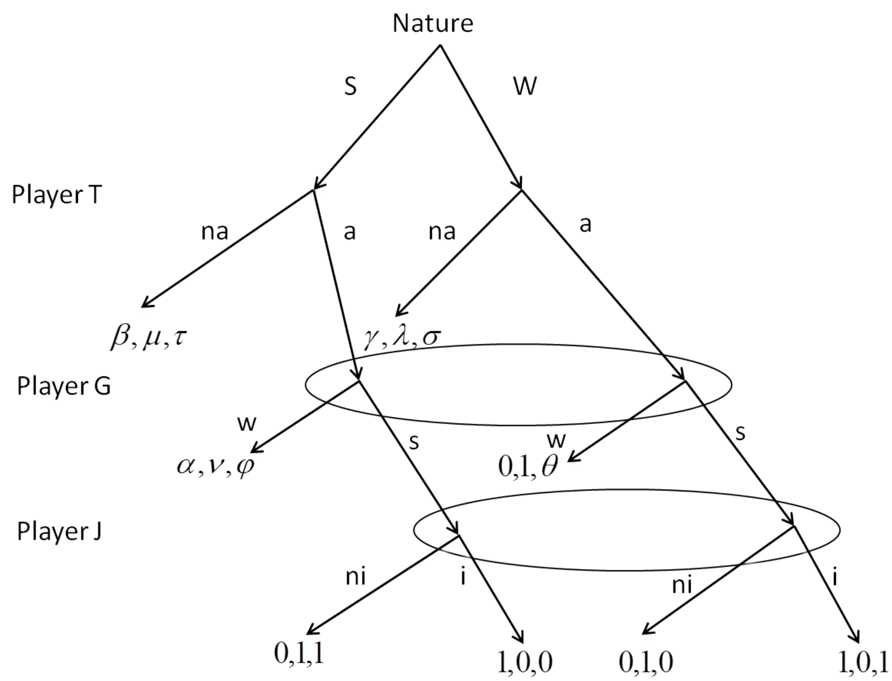
Consider now the game with three players: T and G, as in the Section 2, and the third country, the "Judge", denoted by J. As previously, T can be one of two types: the "weak" (W) or the "strong" (S), p is the common prior belief that T is of the type S. The strategies of T are to attack (a) or not to attack (na). If he chooses not to attack, the game stops. Payoffs of T and G in this case are as in Γ_2. J's outcomes in these cases are denoted by σ if T is of the W-types, and τ otherwise. If T chooses to attack, G chooses one of two strategies: the "weak" reaction (w) or the "strong" reaction (s). If it have chosen the strategy w, the game finishes, and payoffs of T and G are as in Γ_2. J's outcomes in are θ if T is of the W-types, and ϕ otherwise.

If G is attacked and chooses the "strong" reaction s, it destroy the terrorist organization regardless its type. After G have chosen s, J decides abouts its reaction. J prefers the terrorist threat to be removed, but not by "non-proportional" use of force. Namely, it prefers that the "strong" counter terrorist measure would be not used by G against T of the "weak" type W. While observing s strategy used by G, J decides whether to intervene (i) against G, for example, to impose sanctions on G, or not to intervene (ni). The intervention is the worst outcome of G, and the best for T. If G destroyed the terrorist organization by s without afterward intervention, it the most preferable result for G and the worst outcome for T. As for J, its ex-post payoffs depend on the type of T. Its best outcome is to intervene when the type of T is W and not to intervene when the type is S. J's worst outcome is to intervene when the type is S and not to intervene when the type is W.

I denote this game by Γ_3, and it is presented in the Figure 3.1.

In the next proposition, I make an additional assumption: when T chooses not to attack its payoff does not depend on it type. Moreover,to avoid some extreme cases, I assume $p \neq \frac{1}{2}$.

Proposition 3.1. Let $\beta = \gamma$ and $p \neq \frac{1}{2}$. In the game Γ_3, all existing Nash equilibria are of one of the following types:



$$0 < \beta < \alpha < 1$$
$$0 < \lambda, \mu, \nu < 1$$
$$0 < \theta, \varphi, \sigma, \tau < 1$$

Figure 3.1: The three- players game Γ_3

(i) T of both types does not attack, G chooses s with positive probability and J chooses ni with positive probability.

(ii) Both S and W attack with positive probability, G chooses s with positive probability, J mixes i and ni strategies. These equilibria exist only if $\beta \leq \frac{1-\nu}{2}$. In addition, if $p > \frac{1}{2}$, in equilibrium with positive probability of the strategy a G has to choose s with certainty.

The proof appears in Appendix.

Equilibria in Proposition 3.1 are not unique. There are different well-known solution concepts, whose general purpose is to eliminate some Nash equilibria. Among those are solutions due to Selten (1975), to Myerson (1978), to Kreps and Wilson (1982), to Cho and Kreps (1987) and to Banks and Sobel (1987). The basic idea of concepts above is to make some assumptions about beliefs of players, and then to eliminate equilibria which are not compatible with those beliefs. Unfortunately, none of those concepts can resolve the ambiguity of Nash equilibrium results problem. To eliminate some equilibria I introduce a condition of "*compatibility with symmetric mistakes*". It is in spirit of Selten (1975), but imposes a stronger requirement.

In equilibria characterized in part (i) of Proposition 3.1 T does not attack, thus the possibilities of the "strong" reaction by G and of the intervention by J are on an off-equilibrium path. Let us assume that there is a infinitesimally small, but positive, probability that T will attack "by mistake". This is the basic idea in Selten (1975). But my assumption is stronger. I assume that the probability of a "mistake"is symmetric for both types of T.

Formally, let

$$\Pi = \{(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a))| \\ 0 \leq Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a) \leq 1\}$$

be the set of strategy profiles in Γ_3 ($Prob_I(a)$, $I \in \{W, S\}$ is a probability that T of type I chooses the strategy a, $Prob_G(s|a)$ is a probability that G will choose the strategy s, if it was attacked, $Prob_J(i|s,a)$ is a probability that J intervene if G was attacked by T and reacted by s). For $\epsilon > 0$ let

$$\Pi_\epsilon = \{(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a)| \\ \epsilon \leq Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a) \leq 1-\epsilon\}$$

be the set of strategy profiles where each pure strategy is chosen with a probability at least ϵ. That includes profiles where T attacks "by mistake" with some probability. Let $E_\epsilon \subseteq \Pi_\epsilon$ be the set of *ϵ-equilibria*. This is the set of strategies profiles in Π_ϵ, where each strategy is a best reply to strategies of other players, given the constraint that the probability to choose each pure strategy is at least ϵ. Namely,

$$(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a)) \in E_\epsilon \text{ iff}$$

$$Prob_W(a) = \arg \max_{\epsilon \leq Prob_W(a) \leq 1-\epsilon} \{(1 - Prob_W(a))\gamma + Prob_W(a)[Prob_G(s|a) \cdot Prob_J(i|s,a)]\}$$

$$Prob_S(a) = \arg \max_{\epsilon \leq Prob_S(a) \leq 1-\epsilon} \{(1 - Prob_S(a))\beta + Prob_S(a)[(1 - Prob_G(s|a))\alpha + Prob_G(s|a) \cdot Prob_J(i|s,a)]\}$$

$$Prob_G(s|a) = \arg \max_{\epsilon \leq Prob_G(s|a) \leq 1-\epsilon} \{(1 - Prob_G(s|a))[(1 - Prob(W|a))\nu + Prob(W|a)] + \\ + Prob_G(s|a)(1 - P_J(i|s,a))\}$$

$$P_J(i|s,a) = \arg \max_{\epsilon \leq P_J(i|s,a) \leq 1-\epsilon} \{(1 - P_J(i|s,a))(1 - Prob(W|a)) + P_J(i|s,a) Prob(W|a)\}$$

where $Prob(W|a)$ is given by (3).

A Nash equilibrium $(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a)) \in \Pi$ is *compatible with symmetric mistakes* if exists a sequence of $(\epsilon_k)_{k=1}^{\infty}$ which converges to 0, and a sequence of

$((Prob^k_W(a), Prob^k_S(a), Prob^k_G(s|a), Prob^k_J(i|s,a)) \in E_{\epsilon_k})^{\infty}_{k=1}$ which converges to $(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a))$ as $\epsilon_k \to 0$. The next proposition states that for some values of p one can eliminate some equilibria as not compatible with symmetric mistakes. As previously, I assume $\beta = \gamma$.

Proposition 3.2. Let $\beta = \gamma$ and $\beta \leq \frac{1-\nu}{2}$.

(i) Let $p < \frac{1}{2}$. In the game Γ_3 there is no equilibrium compatible with symmetric mistakes where $Prob_W(a) = Prob_S(a) = 0$.

(ii) Let $p > \frac{1}{2}$. In the game Γ_3 there is no equilibrium compatible with symmetric mistakes where $Prob_W(a) > 0$ and $Prob_S(a) > 0$.

The proof appears in appendix.

Remark Note that in the proof of part (ii) of the proposition, the symmetry of mistakes is not used. Thus, it could be shown that there is no perfect equilibrium (in the sense of Selten (1975)) of the form specified in part (ii) of Proposition 3.2. In the proof of part (i) the symmetry of mistakes is crucial.

The next proposition shows existence of an equilibrium compatible with symmetric mistakes for $\beta \leq \frac{1-\nu}{2}$ and $p \neq \frac{1}{2}$.

Proposition 3.3. Let $\beta \leq \frac{1-\nu}{2}$ and $p \neq \frac{1}{2}$. There exists at least one equilibrium compatible with symmetric mistakes.

Let us summarize findings of Theorems 3.1, 3.2 and 3.3. If $\beta > \frac{1-\nu}{2}$ then in all Nash equilibria T does not attack. If $\beta \leq \frac{1-\nu}{2}$ and $p < \frac{1}{2}$, in all equilibria compatible with symmetric mistakes there is positive probability that T of both types will attack, G react with s, and J will choose i. Moreover, such an equilibrium exists. Namely, there is a positive probability that W will attack, and G will use than a "non proportional force" against it. However, it may happen with a positive probability that S attacks, G uses an "appropriate" force s, but is unjustifiably punished by J.

As for the case $\beta \leq \frac{1-\nu}{2}$ and $p > \frac{1}{2}$, in all equilibria compatible with symmetric mistakes, and at least one such an equilibrium exists, T does not attack with certainty.

References

Arce, D. G. and T. Sandler (2007). Terrorist signalling and the value of intelligence. *British Journal of Political Science 37*(04), 573–586.

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

Cho, I. and D. M. Kreps (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 179–221.

Duffy, H. (2015). *The "War on Terror" and the Framework of International Law, 2nd edition*. Cambridge University Press.

Enders, W. and T. Sandler (2011). *The political economy of terrorism*. Cambridge University Press.

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

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

Myerson, R. B. (1978). Refinements of the nash equilibrium concept. *International Journal of Game Theory 7*(2), 73–80.

O'Neill, B. (1994). Game theory models of peace and war. *R. J. Aumann and S. Hart (eds), Handbook of game theory with economic applications 2*, 995–1053, North–Holland, Amsterdam.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4*(1), 25–55.

Wilkinson, P. (2011). *Terrorism versus democracy: The liberal state response.* Taylor & Francis.

Appendix

Proof of Proposition 2.1 Denote by $Prob_I(a)$, $I \in \{W, S\}$ as a probability that T of type I will choose the strategy a. Denote also by $Prob_G(w|a)$ a probability that G will choose the strategy w, given it was attacked. By Figure 2 W prefers the strategy a iff

$$\gamma \leq 1 - Prob_G(w|a) \tag{1}$$

S prefers a to na iff

$$\beta \leq \alpha Prob_G(w|a) \tag{2}$$

Given that G was attacked, the probability it assigns to being T of the type W is

$$Prob(W|a) = \frac{(1-p)Prob_W(a)}{(1-p)Prob_W(a) + pProb_S(a)} \tag{3}$$

G prefers w to s iff

$$(1 - Prob(W|a))\nu + Prob(W|a) \geq 1 - Prob(W|a) \tag{4}$$

Lemma 1. In equilibrium $Prob_W(a) > 0$ implies $Prob_S(a) > 0$, and $Prob_S(a) > 0$ implies $Prob_W(a) > 0$.

Proof Assume to the contrary that $Prob_W(a) > 0$, but $Prob_S(a) = 0$. Then if G is attacked it knows for sure that T is of the type W and G's best reply is w. But then W is better off by deviating to $Prob_W(a) = 0$, contradiction. Similarly, it could be shown that $Prob_S(a) > 0$ implies $Prob_W(a) > 0$.$\square$

Lemma 2. G is not playing any pure strategy in an equilibrium.

Proof If G chooses w with certainty when attacked, then W will not attack for sure, but S will play the strategy a with certainty, contradiction to Lemma 1. If G plays pure s when attacked, W is the best off by choosing a, while S's best choice is to play na. Still, a contradiction to Lemma 1.□

Lemma 3. Let $p \neq \frac{1}{2-\nu}$. Then there is no equilibrium with $Prob_W(a) = Prob_S(a) = 1$.

Proof Suppose an equilibrium with $Prob_W(a) = Prob_S(a) = 1$ exists. By (3), $Prob(W|a) = 1-p$, and by (4), G strongly prefers w iff $p < \frac{1}{2-\nu}$ and strongly prefers s iff $p > \frac{1}{2-\nu}$. Therefore, for $p \neq \frac{1}{2-\nu}$, G is better off by choosing a pure strategy w or s, depending whether p is low or high, contradiction to (2). □

Observe that $0 < Prob_W(a) < 1$ and $0 < Prob_S(a) < 1$ imply equality in (1) and (2), and it holds only if $1 - \gamma = \frac{\beta}{\alpha}$. Thus, and by lemmas 1, 2 and 3, for $1 - \gamma \neq \frac{\beta}{\alpha}$ and $p \neq \frac{1}{2-\nu}$, only three possible equlibrium profiles should be considered: $(Prob_W(a) = 0, Prob_S(a) = 0, 0 < Prob_G(w|a) < 1)$, $(Prob_W(a) = 1, 0 < Prob_S(a) < 1, 0 < Prob_G(w|a) < 1)$ and $(Prob_S(a) = 1, 0 < Prob_W(a) < 1, 0 < Prob_G(w|a) < 1)$.

Case $(Prob_W(a) = 0, Prob_S(a) = 0, 0 < Prob_G(w|a) < 1)$

By (1) and (2) this implies $1 - \gamma \leq Prob_G(w|a) \leq \frac{\beta}{\alpha}$.

Case $(Prob_W(a) = 1, 0 < Prob_S(a) < 1, 0 < Prob_G(w|a) < 1)$

S is indifferent between a and na, when W prefers a. Thus, by (1) and (2) this implies $Prob_G(w|a) = \frac{\beta}{\alpha} \leq 1 - \gamma$.

Given a, G is indifferent between na an a. From (3) and (4), after plugging in $Prob_W(a) = 1$

$$Prob_S(a) = \frac{1-p}{p(1-\nu)},$$

and

$$Prob_S(a) < 1 \Leftrightarrow \frac{1}{2-\nu} < p$$

Case $(Prob_S(a) = 1, 0 < Prob_W(a) < 1, 0 < Prob_G(w|a) < 1)$

W is indifferent between a and na, when S prefers a. Thus, by (1) and (2) this implies $Prob_G(w|a) = 1 - \gamma \geq \frac{\beta}{\alpha}$.

Given a, G is indifferent between na an a. From (3) and (4),

$$Prob_S(a) = \frac{p(1-\nu)}{1-p},$$

and

$$Prob_S(a) < 1 \Leftrightarrow \frac{1}{2-\nu} > p$$

That completes the proof of Proposition 2.1. □

Proof of Proposition 3.1 Denote by $Prob_J(i|s,a)$ the probability that J chooses to intervene given that T have chosen a and G, as a reaction, have chosen s. As in the proof of Proposition 2.1, let $Prob_I(a)$, $I \in \{W, S\}$ be a probability that T of type I chooses the strategy a, and $Prob_G(w|a)$ be a probability that G will choose the strategy w, given it was attacked. Let $Prob_J(i|s,a)$ be probability that J intervene if G was attacked by T and reacted by s. From Figure 3.1, and by assumption $\beta = \gamma$, W prefers a to na iff

$$\beta \leq Prob_G(s|a) \cdot Prob_J(i|s,a), \tag{5}$$

and S prefers a to na iff

$$\beta \leq \alpha(1 - Prob_G(s|a)) + Prob_G(s|a) \cdot Prob_J(i|s,a). \tag{6}$$

The existence of equilibria, where T of both types prefers not to attack, $Prob_J(i|s,a) < 1$ and $Prob_G(s|a) > 1$ is straightforward from (5) and (6).

I proceed now to show conditions for existence of equilibria where T chooses a with positive probability.

Given T attacked G, G and J assign a probability given by (3) that T is of the type W.

While attacked, G prefers the strategy s to w iff

$$\nu(1 - Prob(W|a)) + Prob(W|a) \leq 1 - Prob_J(i|s, a) \tag{7}$$

Given T attacked and G reacted by s, J prefers i iff

$$Prob(W|a) \leq 1 - Prob(W|a) \tag{8}$$

If $Prob_J(i|s, a) = 0$, the only equilibrium is the one with properties of part (i) of the Proposition 3.1. Indeed, the most preferable strategy for G is the only equilibrium possible is s and then the best outcome for T is when na is chosen. To proceed to analyze equlibria with positive probability of an attack, I consider hereafter only equilibria with

$$Prob_J(i|s, a) > 0. \tag{9}$$

Lemma 4. In equilibrium $Prob_W(a) > 0$ implies $Prob_S(a) > 0$.

Proof Assume to the contrary that $Prob_W(a) > 0$, but $Prob_S(a) = 0$. Then if G is attacked it knows for sure that T is of the type W, and takes into account (9), thus G's best reply is w. But then W is better off by deviating to $Prob_W(a) = 0$, contradiction.□

Lemma 5. In equilibrium $Prob_G(s|a) > 0$.

Proof Suppose $Prob_G(s|a) = 0$, namely, G plays w for sure given a. Then the best strategy for W is na, while the best strategy for S is a, contradiction to Lemma 1.□

Lemma 6. In equilibrium $Prob_S(a) > 0$ implies $Prob_W(a) > 0$.

Proof Assume to the contrary that $Prob_S(a) > 0$, but $Prob_W(a) = 0$. Then if G is attacked, G and J know with certainty that T is of the type S. By Lemma 5 there is positive probability that G chooses s, and given it is chooses, the best reply of J, which knows that T's type is S, is ni with certainty. Therefore G's strategy in equilibrium should be s with certainty. But then S is better off by deviating to $Prob_S(a) = 0$, contradiction.□

Lemma 7. There is no equilibrium where $Prob_J(i|s, a) = 1$.

Proof If $Prob_J(i|s, a) = 1$, then the best strategy of G is $Prob_G(s|a) = 0$, contradiction to Lemma 5.□

By (9) and by Lemma 7, I consider only equilibria with

$$0 < Prob_J(i|s, a) < 1 \tag{10}$$

and equality holds in (8). Therefore,

$$Prob(W|a) = \frac{1}{2} \tag{11}$$

Lemma 8. There is no equilibrium where $Prob_W(a) = 1$

Proof Suppose $Prob_W(a) = 1$ in equilibrium. By (3) and (11), $Prob_W(a) = Prob_S(a) = 1$ is possible in equilibrium only if $p = \frac{1}{2}$, but it is assumed that $p \neq \frac{1}{2}$.

Thus, $0 < Prob_S(a) < 1$ (($Prob_S(a) > 0$) by Lemma 4), and it implies equality in (6), but by (5) W strongly prefers na, contradicting $Prob_W(a) = 1$. □

By Lemmas 4-8 and by (10), only two following equilibrium profiles with positive probability of the attack should be considered.

1. $(0 < Prob_W(a) < 1, 0 < Prob_S(a) < 1, Prob_G(s|a) > 0, 0 < (Prob_J(i|s, a) < 1)$

This equilbria implies equality in both (5) and (6), and this is possible only if $Prob_G(s|a) = 1$. Therefore, by (5), $Prob_J(i|s, a) = \beta$, and thus $0 < Prob_J(i|s, a) < 1$ as required, and (11) holds. Thus, any pair of $Prob_W(a)$, $Prob_S(a)$ such that when they are substituted into (3), (11) is satisfied, may be in equilibrium. Since G prefers s, by (7), $\beta \leq \frac{1-\nu}{2}$.

2. $(0 < Prob_W(a) < 1, Prob_S(a) = 1, Prob_G(s|a) > 0, 0 < (Prob_J(i|s,a) < 1)$

If $Prob_G(s|a) = 1$ this case coincides with the previous one. As previously, $Prob_J(i|s,a) = \beta$. Consider $0 < Prob_G(s|a) < 1$. Then equality in (7) holds. By (11) this implies $Prob_J(i|s,a) = \frac{1-\nu}{2}$. Since equality holds in (5),

$$Prob_G(s|a) = \frac{2\beta}{1-\nu}$$

and $0 < Prob_G(s|a) < 1$ iff $\beta < \frac{1-\nu}{2}$. By (11) and (3),

$$Prob_W(a) = \frac{p}{1-p}$$

and $0 < Prob_W(a) < 1$ iff $p < \frac{1}{2}$.

This completes the proof of Proposition 3.1. $\square$.

Proof of Proposition 3.2 Let the strategic profile $(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a))$ be an Nash equilibrium in the game Γ_3. Suppose this profile is compatible with symmetric mistakes, namely, there are sequences $(\epsilon_k)_{k=1}^{\infty} \to 0$ and $((Prob_W^k(a), Prob_S^k(a), Prob_G^k(s|a), Prob_J^k(i|s,a)) \in E_{\epsilon_k})_{k=1}^{\infty}$ which converges to $(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a))$ as $\epsilon_k \to 0$. Recall, $\epsilon_k \le Prob_W^k(a) \le 1-\epsilon_k$, $\epsilon_k \le Prob_S^k(a) \le 1-\epsilon_k$, $\epsilon_k \le Prob_G^k(s|a) \le 1-\epsilon_k$ and $\epsilon_k \le Prob_J^k(i|s,a) \le 1-\epsilon_k$.

Part (i). Consider the case $p < \frac{1}{2}$. Suppose a Nash equilibrium $(Prob_W(a) = 0, Prob_S(a) = 0, 0 < Prob_G(s|a) \le 1, 0 \le Prob_J(i|s,a) < 1)$ is compatible with symmetric mistakes. Let us distinguish two cases.

a. $\beta > Prob_G(s|a) \cdot Prob_J(i|s,a)$ and $\beta > \alpha(1 - Prob_G(s|a)) + Prob_G(s|a) \cdot Prob_J(i|s,a)$. Then there exists K such that for $k > K$, $\beta > Prob_G^k(s|a) \cdot Prob_J^k(i|s,a)$ and $\beta > \alpha(1 - Prob_G^k(s|a)) + Prob_G^k(s|a) \cdot Prob_J^k(i|s,a)$, and by Figure 3.1 T of both types prefers na. Therefore, $Prob_W(a) = Prob_S(a) = \epsilon_k$ (T assigns the minimal allowed probability to its inferior strategy a). By (3),

$$Prob(W|a) = 1 - p,$$

but since $p < \frac{1}{2}$, J should strongly prefer i, contradiction to $Prob_J(i|s,a) < 1$.

b. $\beta > Prob_G(s|a) \cdot Prob_J(i|s,a)$ and $\beta = \alpha(1 - Prob_G(s|a)) + Prob_G(s|a) \cdot Prob_J(i|s,a)$. This implies $0 < P_G(s|a) < 1$, therefore G is indifferent between w and s once attacked. $P_G(s|a) < 1$ also implies $0 < P_J(i|s,a) < 1$, since if $0 = P_J(i|s,a)$, G strongly prefers s. Thus, J is also indifferent between i and ni. Then by (7) and (8),

$$Prob_J(i|s,a) = \frac{1-\nu}{2},$$

and $\beta = \alpha(1 - Prob_G(s|a)) + Prob_G(s|a) \cdot Prob_J(i|s,a)$ implies

$$Prob_G(s|a) = \frac{\alpha - \beta}{\alpha - \frac{1-\nu}{2}}$$

but

$$\frac{\alpha - \beta}{\alpha - \frac{1-\nu}{2}} < 1 \Rightarrow \beta > \frac{1-\nu}{2},$$

contradiction to assumption $\beta \leq \frac{1-\nu}{2}$.

Part (ii). Consider the case $p > \frac{1}{2}$. Suppose the Nash equilibrium $(0 < Prob_W(a) < 1, 0 < Prob_S(a), 0 < Prob_G(s|a) \leq 1, 0 < Prob_J(i|s,a) < 1)$ is compatible with symmetric mistakes. By Proposition 3.1, for $p > \frac{1}{2}$, $Prob_G(s|a) = 1$. Moreover, similar to the proof of Proposition 3.1, $Prob_S(a) = 1$ is impossible in the region $p > \frac{1}{2}$. Therefore, $0 < Prob_W(a) < 1$, $0 < Prob_S(a) < 1$, and there is K such that for $k > K$, $\epsilon_k < Prob^k_W(a) < 1 - \epsilon_k$, $\epsilon_k < Prob^k_S(a) < 1 - \epsilon_k$. Namely, W and S are indifferent between na and a. By Figure 3.1, $\beta = Prob^k_G(s|a) \cdot Prob^k_J(i|s,a)$ and $\beta = \alpha(1 - Prob^k_G(s|a)) + Prob^k_G(s|a) \cdot Prob^k_J(i|s,a)$ and this is contradiction to $Prob^k_G(s|a) \leq 1 - \epsilon_k < 1$. □

Proof of Proposition 3.3 Let $p < \frac{1}{2}$. Consider

$$Prob_W(a) = \frac{p}{1-p}, Prob_S(a) = 1, Prob_G(s|a) = \frac{2\beta}{1-\nu}, Prob_J(i|s,a) = \frac{1-\nu}{2}$$

It is shown in the proof of Proposition 3.1 that it is an Nash equilibrium for $p < \frac{1}{2}$ and $\beta \leq \frac{1-\nu}{2}$. To prove that it is compatible with symmetric mistakes one should show existence of sequences $(\epsilon_k)_{k=1}^{\infty} \to 0$ and
$((Prob_W^k(a), Prob_S^k(a), Prob_G^k(s|a), Prob_J^k(i|s,a)) \in E_{\epsilon_k})_{k=1}^{\infty}$ which converges to
$(Prob_W(a), Prob_S(a), Prob_G(s|a), Prob_J(i|s,a))$ as $\epsilon_k \to 0$.

For ϵ_k sufficiently low one can define

$$\epsilon_k < Prob_W(a) = Prob_W^k(a) < 1 - \epsilon_k,$$

$$\epsilon_k < Prob_G(s|a) = Prob_G^k(s|a) < 1 - \epsilon_k,$$

and

$$\epsilon_k < Prob_J(i|s,a) = Prob_J^k(i|s,a) < 1 - \epsilon_k.$$

Let $Prob_S^k(a) = 1 - \epsilon_k$. Since from $0 < Prob_W(a) < 1$, $\beta = Prob_G(s|a) \cdot Prob_J(i|s,a)$ and $Prob_G(s|a) = Prob_G^k(s|a) < 1$, then $\beta < \alpha(1 - Prob_G^k(s|a)) + Prob_G^k(s|a) \cdot Prob_J^k(i|s,a)$. Therefore, $Prob_S^k(a) = 1 - \epsilon^k$ is the best strategy for S, and it converges to $Prob_S(a) = 1$.

Next consider the case $p > \frac{1}{2}$. Let us consider a Nash equilibrium

$$Prob_W(a) = Prob_S(a) = 0, Prob_G(s|a) = 1, Prob_J(i|s,a) = 0$$

Consider

$$Prob_W^k(a) = Prob_S^k(a) = Prob_J(i|s,a) = \epsilon_k, Prob_G(s|a) = 1 - \epsilon_k$$

It is sufficient to show that this strategy profile belongs to E_{ϵ_k} for sufficient low ϵ_k.

$\beta > Prob_G^k(s|a) \cdot Prob_J^k(i|s,a)$ and $\beta > \alpha(1 - Prob_G^k(s|a)) + Prob_G^k(s|a) \cdot Prob_J^k(i|s,a)$ for sufficient low ϵ_k, therefore $Prob_W^k(a) = Prob_S^k(a) = \epsilon_k$ are the best strategies for W and S. By (3),

$$Prob(W|a) = 1 - p < \frac{1}{2}$$

and therefore $Prob_J(i|s,a) = \epsilon_k$ is the best reply for J. Then, for ϵ_k sufficiently low, $p\nu + 1 - p < 1 - \epsilon_k = 1 - Prob_J^k(i|s,a)$, and G's best strategy is $Prob_G(s|a) = 1 - \epsilon_k$. $\square$