Undergraduate Formal IR Game Theory Notes

Hankyeul Yang

November 19, 2025

Introduction

The purpose of the notes provided here is to help undergraduate students read various International Relations articles using formal models. More notes to be added.

1 Revealing Preferences (Lewis and Schultz 2003)

A, B: two states
SQ: outcome when A does not make a challenge.
V_A: the value that A places on getting the good without a fight
S_A: A's payoff from the status quo
p_F: probability of A standing firm at its last node
p_R: probability of B resisting A's challenge
p_C: probability of A making a challenge

Assumption:
1. Good belongs to B.
2. Audience cost is not necessarily less than zero. $a \in \mathcal{R}$
3. $\overline{W}_A$, $\overline{W}_B$, $\overline{a}$ are common knowledge
4. Disturbance terms are known only by the appropriate state

Order of play:
1. A decides whether or not to challenge B.
2. If A does not make a challenge the status quo prevails.

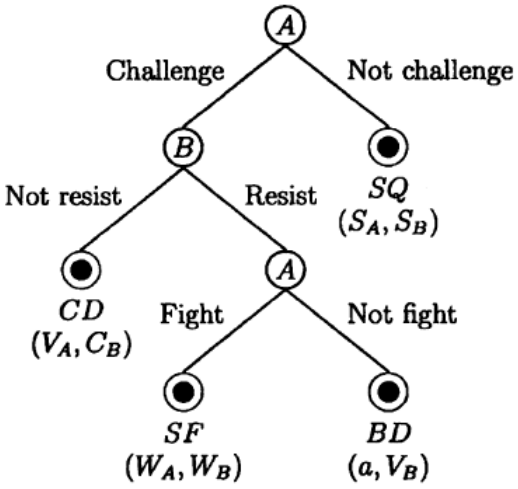

Fig. 1 Simple crisis bargaining game.

p. 350

$$W_A = \overline{W}_A + \epsilon_A$$
$$W_B = \overline{W}_B + \epsilon_B$$
$$a = \overline{a} + \epsilon_a$$
$$\epsilon_A \sim \mathcal{N}(0, \sigma^2)$$
$$\epsilon_B \sim \mathcal{N}(0, \sigma^2)$$
$$\epsilon_a \sim \mathcal{N}(0, \sigma^2)$$
$$p_F \equiv Pr(W_A > a)$$

At its final node, A fights iff $W_A > a$
$p_F \equiv Pr(W_A > a)$
B's expected utility for not resisting given the posterior belief p_F is simply C_B
B's expected utility for resisting given the posterior belief p_F is $p_F W_B + (1 - p_F)V_B$
Thus, B resists if $p_F W_B + (1 - p_F)V_B > C_B$

$$p_F W_B + (1 - p_F)V_B > C_B$$
$$\Rightarrow W_B > \frac{C_B - (1 - p_F)V_B}{p_F}$$
$$\Rightarrow \overline{W}_B + \epsilon_B > \frac{C_B - (1 - p_F)V_B}{p_F}$$
$$\epsilon_B > \frac{C_B - (1 - p_F)V_B - p_F\overline{W}_B}{p_F}$$
$$\Rightarrow p_R \equiv Pr\left(\epsilon_B > \frac{C_B - (1 - p_F)V_B - p_F\overline{W}_B}{p_F}\right) = 1 - Pr\left(\epsilon_B < \frac{C_B - (1 - p_F)V_B - p_F\overline{W}_B}{p_F}\right)$$
$$= 1 - \mathbf{\Phi}\left[\frac{C_B - (1 - p_F)V_B - p_F\overline{W}_B}{p_F\sigma}\right]$$
$$= \mathbf{\Phi}\left[\frac{p_F\overline{W}_B + (1 - p_F)V_B - C_B}{p_F\sigma}\right]$$

by the symmetrical nature of the normal distribution.[1]

Given p_R, the expected value of making a challenge for an A of type (a, W_A) is $EU_A(CH) = p_R \max(a, W_A) + (1 - p_R)V_A$
The expected value of status quo for A is simply S_A
Thus, A challenges if

$$EU_A(CH) > EU_A(SQ)$$
$$\Rightarrow p_R \max(a, W_A) + (1 - p_R)V_A > S_A$$
$$\Rightarrow \max(a, W_A) > \frac{S_A - (1 - p_R)V_A}{p_R} \equiv c^*$$

[1] You can check in R: 1-pnorm(0.2/7) gives the same result as pnorm(-0.2/7)

$$
\begin{aligned}
p_C &\equiv Pr\left(\max(a, W_A) > \frac{S_A - (1-p_R)V_A}{p_R}\right) \\
&= 1 - Pr(W_A < c^*)Pr(a < c^*) \\
&= 1 - Pr(\overline{W}_A + \epsilon_A < c^*)Pr(\overline{a} + \epsilon_a < c^*) \\
&= 1 - Pr(\epsilon_A < c^* - \overline{W}_A)Pr(\epsilon_a < c^* - \overline{a}) \\
&\qquad \text{since } \epsilon_i \sim N(0, \sigma), \\
&= 1 - \Phi\left(\frac{c^* - \overline{W}_A}{\sigma}\right)\left(\frac{c^* - \overline{a}}{\sigma}\right)
\end{aligned}
$$

Note that p_C is equivalent to one minus the probability that both a and W_A are less than c^*

$$
\begin{aligned}
p_F &= Pr[W_A > a | \max(a, W_A) > c^*] \\
&= Pr[W_A > a \cap \max(a, W_A]/Pr[\max(a, W_A) > c^*] \\
&= Pr[W_A > a \cap \max(a, W_A) > c^*]/p_C \\
&= Pr[W_A - a > 0 \cap W_A > c^*]/p_C \\
&\Rightarrow p_F = \Phi_2\left(\frac{\overline{W}_A - \overline{a}}{\sigma\sqrt{2}}, \frac{\overline{W}_A - c^*}{\sigma}, \frac{1}{\sqrt{2}}\right) \Big/ p_C
\end{aligned}
$$

For the derivation in the last line, refer to the detailed derivation below.

For calculating joint normal distribution refer to section 5.3.2 in the following[2]

Two random variables X and Y are said to have a bivariate normal distribution with parameters μ_X, σ_X^2, μ_Y, σ_Y^2, and ρ if their joint PDF is given by

$$
f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}
$$

where $\mu_X, \mu_Y \in \mathcal{R}$, $\sigma_X, \sigma_Y > 0$ and $\rho \in (-1, 1)$ are all constants.

$$
\begin{aligned}
\Delta &\equiv W_A - a \\
W_A &\sim \mathcal{N}(\overline{W}_A, \sigma^2) \\
a &\sim \mathcal{N}(\overline{a}, \sigma^2) \\
\therefore \Delta &\sim \mathcal{N}(\overline{W}_A - \overline{a}, 2\sigma^2) \\
Cov(\Delta, W_A) &= Cov(W_A - a, W_A) = Var(W_A) - Cov(W_A, a) = Var(W_A) = \sigma^2 \\
\rho_{\Delta, W_A} &\equiv \frac{Cov(\Delta, W_A)}{\sigma_\Delta \sigma_{W_A}} \\
&= \frac{\sigma^2}{\sqrt{2}\sigma\sigma} = \frac{1}{\sqrt{2}}
\end{aligned}
$$

[2] https://www.probabilitycourse.com/chapter5/5_3_2_bivariate_normal_dist.php

2 Decentralization, Repression, and Gambling for Unity (Gibilisco 2021)

C: Center
P: Periphery
$t \in N$
$g^t \in N_0$
$\delta \in (0,1)$
$F(g^t) \in [0,1]$
$\lim_{g\to\infty} F(g) = p$
$r^t \in \{\emptyset, 0, 1\}$
$\pi_i^j \geq 0$: payoff for i per period when j controls the territory
$\pi_C^P = 0$: Center's benefit under Periphery control
$\psi > 0$: cost to Center if the Periphery successfully mobilizes a secessionist movement (note that this cost is NOT incurred should the Center grant independence)
κ_C: repression cost for Center
κ_P: mobilization cost for Periphery
$V_i^\sigma(g)$: i's continuation value from beginning the game at grievance g when both actors subsequently play according to profile σ
$\tilde{V}_C(g)$: Center's continuation value from beginning at grievance g and continuing to neither repress nor grant independence in all future periods while the Periphery mobilizes if and only if $g > g^-$; in other words, this is the Center's expected utility from *gambling for unity* at grievance g, that is, from tolerating secessionist mobilization until grievances reach peaceful levels.
$\overline{V}_i$: i's continuation value after a history in which the Periphery has won control of the territory; $\overline{V}_C = 0$ and $\overline{V}_P = \frac{\pi_P^P}{1-\delta}$

$$g^{t+1} = \begin{cases} g^t + 1 & \text{if } r^t = 1 \\ \max\{g^t - 1, 0\} & \text{otherwise} \end{cases}$$

Assumption 1: Periphery values independence

$$
\begin{aligned}
EU_P(\textit{mobilize once}|F(g)=p) &> EU_P(\textit{never mobilize}) \\
\Rightarrow p\frac{\pi_P^P}{1-\delta} + (1-p)\frac{\pi_P^C}{1-\delta} - \kappa_P &> \frac{\pi_P^C}{1-\delta} \\
\Rightarrow p\frac{\pi_P^P - \pi_P^C}{1-\delta} &> \kappa_P \\
\Rightarrow \pi_P^P - \pi_P^C &> \frac{\kappa_P(1-\delta)}{p}
\end{aligned}
$$

Assumption 2: Secession is costly

$$EU_C(P \text{ mobilizes every period}; r = 0)$$
$$= \Big\{(1-p)\pi_C^C - p\psi\Big\} + (1-p)\delta\Big\{(1-p)\pi_C^C - p\psi\Big\} + (1-p)^2\delta^2\Big\{(1-p)\pi_C^C - p\psi\Big\}$$
$$= \frac{(1-p)\pi_C^C - p\psi}{1-(1-p)\delta}$$
$$EU_C(C \text{ grants independence}) = 0$$
$$EU_C(C \text{ represses every period}) = \frac{\pi_C^C - \kappa_C}{1-\delta}$$

Note that we have to put the powers on $(1-p)$ as well for $EU_C(P \text{ mobilizes every period}; r = 0)$ since we are assuming that the mobilization was not successful for two periods, three periods, etc.

To formalize the assumption that secession is costly we now let the first expected utility be smaller than the second expected utility to formalize the notion that the Center would rather grant independence or repress every period

$$EU_C(P \text{ mobilizes every period}; r = 0) < \max\{EU_C(C \text{ grants independence}), EU_C(C \text{ represses every period})\}$$
$$\Rightarrow \frac{(1-p)\pi_C^C - p\psi}{1-(1-p)\delta} < \max\left\{0, \frac{\pi_C^C - \kappa_C}{1-\delta}\right\}$$
$$\Rightarrow (1-p)\pi_C^C - p\psi < \max\{0, \frac{(1-(1-p)\delta)(\pi_C^C - \kappa_C)}{1-\delta}\}$$
$$\Rightarrow -p\psi < \max\left\{-(1-p)\pi_C^C, \frac{(1-(1-p)\delta)(\pi_C^C - \kappa_C)}{1-\delta} - \frac{((1-p)\pi_C^C)(1-\delta)}{1-\delta}\right\}$$
$$\Rightarrow -p\psi < \max\left\{-(1-p)\pi_C^C, \frac{(1-\delta+\delta p)(\pi_C^C - \kappa_C) - (\pi_C^C - p\pi_C^C)(1-\delta)}{1-\delta}\right\}$$
$$\Rightarrow -p\psi < \max\left\{-(1-p)\pi_C^C, \frac{(\pi_C^C - \kappa_C - \delta\pi_C^C + \delta\kappa_C + \delta p\pi_C^C - \delta p\kappa_C) - (\pi_C^C - \delta\pi_C^C - p\pi_C^C + \delta p\pi_C^C)}{1-\delta}\right\}$$
$$\Rightarrow -p\psi < \max\left\{-(1-p)\pi_C^C, \frac{-\kappa_C + \delta\kappa_C - \delta p\kappa_C + p\pi_C^C}{1-\delta}\right\}$$
$$\Rightarrow -p\psi < \max\left\{-(1-p)\pi_C^C, \frac{-(1-\delta)\kappa_C - p(\delta\pi_C^C - \kappa_C)}{1-\delta}\right\}$$
$$\therefore \psi > \min\left\{\frac{\pi_C^C(1-p)}{p}, \frac{(1-\delta)\kappa_C - p(\pi_C^C - \delta\kappa_C)}{p(1-\delta)}\right\}$$

Small Grievances

For a given level of grievance,

$$EU_P(mobilize) = -\kappa_P + F(g)\frac{\pi_P^P}{1-\delta} + (1-F(g))\pi_P^C + \delta\Big\{(1-F(g))(V_P^\sigma(\max\{g-1,0\}))\Big\}$$
$$EU_P(\neg mobilize) = \pi_P^C + \delta V_P^\sigma(\max\{g-1,0\})$$

Thus, to find the level of grievance at which the Periphery would not have the incentive to mobilize,

$$EU_P(\neg mobilize) \geq EU_P(mobilize)$$
$$\Rightarrow \pi_P^C + \delta V_P^\sigma(\max\{g-1,0\}) \geq -\kappa_P + F(g)\frac{\pi_P^P}{1-\delta} + (1-F(g))\pi_P^C + \delta\Big\{(1-F(g))V_P^\sigma(\max\{g-1,0\})\Big\}$$
$$\Rightarrow \kappa_P \geq -F(g)\pi_P^C + F(g)\frac{\pi_P^P}{1-\delta} - \delta F(g)V_P^\sigma(\max\{g-1,0\})$$
$$\therefore \kappa_P \geq F(g)\left[\frac{\pi_P^P}{1-\delta} - \pi_P^C - \delta V_P^\sigma(\max\{g-1,0\})\right]$$

which is what we have for equation (1) on pg. 1358.

Note that now we can obtain the bound for small grievances

$$-F(g)\pi_P^C - \delta F(g)V_P^\sigma(\max\{g-1,0\})$$
$$= -F(g)\Big\{\pi_P^C + \delta V_P^\sigma(\max\{g-1,0\})\Big\}$$
$$\geq -F(g)\Big\{\pi_P^C + \delta\pi_P^C + \delta^2\pi_P^C + ...\Big\} \quad \because P \text{ is guaranteed } \pi_P^C \text{ every period}$$
$$= -F(g)\frac{\pi_P^C}{1-\delta}$$
$$\therefore g^- \equiv \max\left\{g \in N_0 | \kappa_P \geq F(g)\frac{\pi_P^P - \pi_P^C}{1-\delta}\right\}$$

Moderate Grievances

To repeat, $\tilde{V}_C(g)$ refers to the Center's continuation value from beginning at grievance g and continuing to neither repress nor grant independence in all future periods while the Periphery mobilizes if and only if $g > g^-$; in other words, this is the Center's expected utility from *gambling for unity* at grievance g, that is, from tolerating secessionist mobilization until grievances reach peaceful levels.

$$\tilde{V}_C(g) = \begin{cases} \frac{\pi_C^C}{1-\delta} & \text{if} \quad g \leq g^- \\ -F(g)\psi + (1-F(g))\pi_C^C + \delta\Big\{(1-F(g))\tilde{V}_C(g-1)\Big\} \end{cases}$$

The intuition about $\tilde{V}_C(g)$ is that the expected utility is strictly decreasing in the current level of grievance when $g \geq g^-$ because larger grievances imply that the Center will need to wait additional periods before a lasting peace emerges, thereby raising the risk successful mobilization in the gambling for unity dynamic.

Note that Assumption 2 implies that the Center would prefer to either grant independence or repress every period rather than to have the Periphery mobilize every period. Thus,

$$\lim_{g\to\infty} \tilde{V}_C(g) < \max\left\{\frac{\pi_C^C - \kappa_C}{1-\delta}, 0\right\}$$

$$\therefore \exists\ g^+ \in N_0 \text{ s.t. } g < g^+ \iff \tilde{V}_C(g) > \max\left\{\frac{\pi_C^C - \kappa_C}{1-\delta}, 0\right\}$$

Dynamic Payoffs

$U_C^\sigma(r;g)$ denotes the Center's dynamic payoffs from choosing $r \in \{\emptyset, 0, 1\}$ given greivance g when actors subsequently paaly according to profile σ

$$U_C^\sigma(r;g) = \begin{cases} 0 \quad \text{if} \quad r = \emptyset \\ \pi_C^C - \kappa_C + \delta V_C^\sigma(g+1) \quad \text{if} \quad r = 1 \\ -\sigma_P(g)F(g)\psi + \sigma_P(g)(1-F(g))(\pi_C^C + \delta V_C^\sigma(\max\{g-1,0\})) \\ +(1-\sigma_P(g))(\pi_C^C + \delta V_C^\sigma(\max\{g-1,0\})) \quad \text{if} \quad r = 0 \end{cases}$$

$$= \begin{cases} 0 \quad \text{if} \quad r = \emptyset \\ \pi_C^C - \kappa_C + \delta V_C^\sigma(g+1) \quad \text{if} \quad r = 1 \\ -\sigma_P(g)F(g)\psi + (1-\sigma_P(g)F(g))(\pi_C^C + \delta V_C^\sigma(\max\{g-1,0\})) \quad \text{if} \quad r = 0 \end{cases}$$

Similarly, $U_P^\sigma(m;g)$ denotes the Periphery's dynamic payoffs from choosing $min\{0,1\}$ given greivance g when actors subsequently paly according to profil σ

$$U\sigma_P(m;g) = \begin{cases} -\kappa + F(g)\overline{V}_P + (1-F(g))(\pi_P^C + \delta V_P^\sigma(\max\{g-1,0\})) \quad \text{if} \quad m = 1 \\ \pi_P^C + \delta V_P^\sigma(\max\{g-1,0\}) \quad \text{if} \quad m = 0 \end{cases}$$

3 International Crises and Domestic Politics (Smith 1998)

A decides whether to attack, B decides whether to retaliate, C decides whether to intervene on B's behalf if B retaliates.
Four players: A, B, C, and C's domestic audience.
1: value of the prize
$m \in M$: costless message from C indicating her foreign policy
$\Theta = (\theta_a, \theta_b, \theta_c)$: competence or type of the nations
$0 \leq \theta_a \leq 1$: competence of A
$0 \leq \theta_b \leq 1$: competence of B
$0 \leq \theta_c \leq 1$: competence of C
$\theta_a^*(m)$: the type that is indifferent about whether to attack
$\theta_b^*(m)$: the type that is indifferent about whether to retaliate
$\theta_c^*(m)$: the type that is indifferent about whether to intervene
$\mu_i(\theta_i)$: the prior probability density over θ_i (the **beliefs** of the other players about the competence of the leader in nation i) where $\mu_i(\theta_i) = 1$ for $\theta_i \in [0, 1]$
$\alpha(m)$: the probability that A attacks after observing message m
$\beta(m) = \int_{\theta_b^*(m)}^{1} \mu_b(\theta_b|m)d\theta_b = 1 - \theta_b^*(m)$: the probability that B retaliates after observing message m
$\gamma(m) = \int_{\theta_c^*(m)}^{1} \mu_c(\theta_c|m)d\theta_c = 1 - \theta_c^*(m)$: the probability that C intervenes given message m
$\Theta = (\theta_a, \theta_b, \theta_c)$
$q(\Theta) = \frac{\theta_b - \theta_a}{6} + 0.55$: the probability that B wins a bilateral war
$p(\Theta) = \frac{\theta_b + \theta_c - \theta_a}{6} + 0.6$: the probability that B wins a multilateral war
$\mu_c(\theta_c)$: voters' prior beliefs about C (assumed to be uniform)
$\overline{p}(\Theta) = \int_{\theta_b^*(m)}^{1} \int_{\theta_a^*(m)}^{1} p(\Theta)d\theta_a d\theta_b$: the average probability of victory if C intervenes
$\overline{q}(\Theta) = \int_{\theta_b^*(m)}^{1} \int_{\theta_a^*(m)}^{1} q(\Theta)d\theta_a d\theta_b$: the average probabiliy of victory for B in a bilateral war
k_a: cost of fighting for A
k_b: cost of fighting for B
k_c: cost of fighting for C
z: international outcome, namely one of multilateral war, bilateral war, acquiescence, status quo
$\mathbb{E}[\theta_c|m, z]$: voters' belief of the expected competence of C
$\Phi(\mathbb{E}[\theta_c|m, z], bias)$: the probability that the citizens will reelect the incumbent given the beliefs

Note that since we are assuming that $\theta_i \sim \mathcal{U}[0,1]$, we can use $\theta_a^*(m)$ to denote both the type that is indifferent between attacking and attacking; and the CDF of the type that would not attack. ie $\gamma(m) = \int_{\theta_c^*(m)}^{1} \mu_c(\theta_c|m)d\theta_c = 1 - \int_0^{\theta_c^*(m)} \mu_c(\theta_c|m)d\theta_c = 1 - \theta_c^*(m)$. Analogous interpretations follow for $\theta_b^*(m)$ and $\theta_c^*(m)$.

Order of the game:
1. C announces a foreign policy message, $m \in M$
2. Having observed this message, A chooses whether to attack $(att, \neg att)$
3. If A attacks, then B chooses whether to retaliate $(ret, \neg ret)$
4. If B retaliates, then C chooses whether to intervene $(int, \neg int)$

Assumptions regarding competence and probability of B winning (should make intuitive sense)

$$\frac{dq(\Theta)}{d\theta_a} < 0$$
$$\frac{dq(\Theta)}{d\theta_b} > 0$$
$$\frac{dq(\Theta)}{d\theta_c} = 0$$
$$\frac{dp(\Theta)}{d\theta_a} < 0$$
$$\frac{dp(\Theta)}{d\theta_b} > 0$$
$$\frac{dp(\Theta)}{d\theta_c} > 0$$

$\mu_i(\theta_i)$: the beliefs of the other players about the competence of the leader in nation i (assumption: uniform distribution over unit interval) For example, $\mu_a(\theta_a)$ would be what B and C think about A's competence
$\mu_a(\theta_a|att)$: posterior beliefs about A's type, given that it attacks
$\mu_b(\theta_b|ret)$: posterior beliefs about B's type, given that it retaliates
$\mu_c(\theta_c|m)$: posterior distribution of θ_c, given the message m

$s_a : \theta_a \times M \to [0,1]$ A's strategy $(att, \neg att)$
$s_b : \theta_b \times M \to [0,1]$ B's strategy $(ret, \neg ret)$
(σ_c, s_c) where $\sigma_c : \theta_c \times M \to [0,1]$ and $s_c : \theta_c \times M \to [0,1]$ C's strategy
$\sigma_c(m, \theta_c)$: the probability that type θ_c sends message m
$s_a(\theta_a, m)$: the probability that type θ_a attacks having observed the message m
$s_b(\theta_b, m)$: the probability that type θ_b attacks having observed the message m
$s_c(\theta_c, m)$: the probability that type θ_c intervenes having observed the message m
$\Phi(\mathbb{E}[\theta_c|m, z], bias)$: the probability that the voters reelect C
z: international outcome
$\Psi > 0$: payoff for leadership of C of being reelected following the international crisis
$Z = \{\text{MUWAR, BIWAR, ACQ, SQ}\}$: set of international outcomes (multilateral war, bilateral war, acquiescence, status quo)

Proposition 1

For any beliefs about C's type $\mu_c(\theta_c|m)$, the behavior of nations A, B, and C can be characterized by a unique triple: $(\theta_a^*(m), \theta_b^*(m), \theta_c^*(m))$. A only attacks if its type is greater than $\theta_a^*(m)$, B only retaliates if its type is greater than $\theta_b^*(m)$, and C only intervenes if its type is greater than $\theta_c^*(m)$. Having observed m, the probability that A attacks is $\alpha(m) = 1 - \theta_a^*(m)$, the probability that B retaliates is $\beta(m) = 1 - \theta_b^*(m)$, and the probability that C intervenes is $\gamma(m) = \int_{\theta^*c}^{1}(m)\mu_c(\theta_c|m)d\theta_c$

C's expected utility for intervention

$$U_c(int|\theta_c, m) = \int_0^1 \int_0^1 \mu_a(\theta_a|att)\mu_b(\theta_b|ret)p(\Theta)d\theta_a d\theta_b - k_c + \Psi\Phi(\mathbb{E}[\theta_c|m, \text{MUWAR}])$$

Remember that the μ component is just the posterior belief and is analogous to the posterior probability that we use in simple signaling game. (Think of how we multiply the posterior probability to the payoff at each of the nodes for calculating the expected payoff)

C's expected utility for not intervening

$$U_c(\neg int|\theta_c, m) = \int_0^1 \int_0^1 \mu_a(\theta_a|att)\mu_b(\theta_b|ret)q(\Theta)d\theta_a d\theta_b + \Psi\Phi(\mathbb{E}[\theta_c|m, \text{BIWAR}])$$

B's expected utility for retaliating

$$\begin{aligned} U_b(ret|\theta_b, m) &= \int_0^1 \mu_a(\theta_a|att)\bigg(\int_{\theta_c^*(m)}^1 \mu_c(\theta_c|m)p(\Theta)d\theta_c + \int_0^{\theta_c^*(m)} \mu_c(\theta_c|m)d\theta_c q(\Theta)\bigg)d\theta_a - k_b \\ &= \int_0^1 \mu_a(\theta_a|att)\bigg(\int_{\theta_c^*(m)}^1 \mu_c(\theta_c|m)p(\Theta)d\theta_c + (1-\gamma(m))q(\Theta)\bigg)d\theta_a - k_b \end{aligned}$$

A's expected utility for attacking

$$\begin{aligned} U_a(att|\theta_a, m) &= \bigg(1 \times \int_0^{\theta_b^*(m)} \mu(\theta_b|m)d\theta_b\bigg) + \bigg(\int_{\theta_c^*(m)}^1 \int_{\theta_b^*(m)}^1 \mu_c(\theta_c|m)\mu_b(\theta_b|m)(p(\Theta) - k_a)d\theta_b d\theta_c\bigg) + \\ &\qquad \bigg(\int_0^{\theta_c^*(m)} \int_{\theta_b^*(m)}^1 \mu_c(\theta_c|m)\mu_b(\theta_b|m)(q(\Theta) - k_a)d\theta_b d\theta_c\bigg) \\ &= 1 - \beta(m) + \bigg(\int_{\theta_c^*(m)}^1 \int_{\theta_b^*(m)}^1 \mu_c(\theta_c|m)\mu_b(\theta_b|m)(p(\Theta) - k_a)d\theta_b d\theta_c\bigg) + \\ &\qquad \bigg((1-\gamma(m))\int_{\theta_b^*(m)}^1 \mu_b(\theta_b|m)(q(\Theta) - k_a)d\theta_b\bigg) \end{aligned}$$

4 Debs and Monteiro: Known Unknowns (2014)

Players:
T: target
D: deterrer
$k > 0$: investment cost
$I_t = 1$: T makes investment in period t
$I_t = 0$: T makes no investment in period t
$s_t = 1$: the signal that T made an investment in period t
$p_s \in [0, 1]$: the probability with which $s_t = 1$
$M_t \in \{0, 1\}$: T's current military capabilities with $M_t = 1$ if and only if T has acquired additional military capabilities.
$w_T(1)$: T's war payoff with additional military capabilities that it has acquired
$w_T(0)$: T's war payoff with no additional military capabilities
$w_T(M_t) + w_D(M_t) < 1$: war is inefficient
z_t: D's offering of a share of the pie, keeping $1 - z_t$ for itself and conceding z_t to T

Let's think about some of the conditions:
Condition 1

$$\delta[w_T(1) - w_T(0)] \leq k$$
$$\text{effect of militarization} \leq \text{cost of investment}$$

Condition 2

$$\delta[w_T(1) - w_T(0)] \leq 1 - w_T(0) - w_D(0)$$
$$\text{effect of militarization} \leq \text{cost of preventive war}$$

Condition 3

$$(1 - p_s)\delta[w_T(1) - w_T(0)] \leq k$$
$$(\text{probability of } s_t = 0) \times \text{effect of militarization} \leq \text{cost of investment}$$

Intuitively, this is the effect of militarization weighted by the probability that the signal is ambiguous being smaller than the cost of investment. The higher the p_s the expression on LHS would be smaller than the cost of investment, meaning that the effect of militarization isn't worth it.

Incomplete Information (Two Periods)

Proposition 1: *In period 2, there is always peace, where D offers $z_2^* = w_T(M_2)$ and T accepts any $z_2 \geq w_T(M_2)$*

Proposition 2: *In period 1, there is always peace if the effect of militarization is smaller than the cost of a preventive war or smaller than the cost of the investment.*

Note first that in any equilibrium T accepts any offer $z_1 \geq w_T(0)$

4.1 D's Decision to Offer Peaceful Settlement or Launch Preventive War when the Effect of Militarization is Smaller than the Cost of Preventive War

Now consider whether D would offer $z_1 = w_T(0)$

$$\begin{gathered} EU_D(\text{preventive war}) = w_D(0) + \delta(1 - w_T(0)) \\ EU_D(\text{offering } z_1 = w_T(0)) = 1 - w_T(0) + \text{second period payoff} \\ \geq 1 - w_T(0) + \delta(1 - w_T(1)) \end{gathered}$$

Here $w_D(0)$ denotes D's payoff from the war in the first period while $\delta(1 - w_T(0))$ is the payoff in the second period from offering $z_2 = w_T(0)$

Now consider the condition that the effect of militarization is smaller than the cost of preventive war. Then D prefers peace to preventive war if

$$\begin{gathered} \delta[w_T(1) - w_T(0)] < 1 - w_T(0) - w_D(0) \\ \iff EU_D(\text{offering } z_1 = w_T(0)) > EU_D(\text{preventive war}) \\ 1 - w_T(0) + \delta(1 - w_T(1)) > w_D(0) + \delta(1 - w_T(0)) \\ 1 - w_T(0) - w_D(0) + \delta(w_T(0) - w_T(1)) > 0 \end{gathered}$$

Note that the final line is just a rearrangement of $\delta[w_T(1) - w_T(0)] \leq 1 - w_T(0) - w_D(0)$, i.e. the condition that the effect of militarization is smaller than the cost of preventive war.

4.2 T's Decision Whether to Invest in Military Capabilities when the Effect of Militarization is Smaller than the Cost of Investment

$$\begin{gathered} EU_T(\text{Invest}) \leq -k + w_T(0) + \delta w_T(1) \\ EU_T(\neg\text{Invest}) = w_T(0) + \delta w_T(0) \end{gathered}$$

Now let's see what the condition of effect of militarization being smaller than the cost of investment is

$$\begin{gathered} \delta[w_T(1) - w_T(0)] \leq k \\ \iff w_T(0) + \delta w_T(0) \geq -k + w_T(0) + \delta w_T(1) \\ k \geq \delta[w_T(1) - w_T(0)] \end{gathered}$$

Proposition 3: *Consider period 1 and assume that the effect of militarization is greater than the cost of a preventive war and greater than the cost of the investment.*
1. If the signal is sufficiently informative, that is,

$$(1-p_s)\delta[w_T(1)-w_T(0)] \leq k$$

then peace prevails.

2. If the signal is not sufficiently informative, that is $(1-p_s)\delta[w_T(1)-w_T(0)] > k$, then T invests with the following probability

$$q^* = \frac{1}{p_s + (1-p_s)\frac{\delta[w_T(1)-w_T(0)]}{1-w_T(0)-w_D(0)}}$$

After $s_1 = 0$, D offers $z_1^* = w_T(0)$ with the following probability:

$$r^* = \frac{k}{(1-p_s)\delta[w_T(1)-w_T(0)]}$$

and declares war with probability $1-r^*$. After $s_1 = 1$, D declares war. T accepts $z_1 \geq w_T(0)$.

Indifference conditions

$$(1+\delta)w_T(0) = -k + (1+\delta)w_T(0) + (1-p_s)r^*\delta(w_T(1)-w_T(0))$$
$$EU_T(\neg\text{Invest}) = EU_T(\text{Invest})$$
$$w_D(0) + \delta(1-w_T(0)) = (1+\delta)(1-w_T(0)) - \frac{q^*(1-p_s)}{1-q^*p_s}\delta(w_T(1)-w_T(0))$$
$$EU_D(\text{Declares War}) = EU_D(\neg\text{Declares War})$$

Think about why we have such indifference conditions. The first indifference condition LHS should be obvious. The first indifference condition RHS consists of the (i) cost of investment; (ii) the minimum peace payoff guaranteed; and (iii) additional expected payoff when the investment is not detected and D offers $z_1^* = w_T(0)$, thereby allowing T's additional acquisition of military capabilities to be materialized. The second indifference condition LHS consists of first period war payoff and second period peace payoff.[3] The second indifference condition RHS consists of (i) the maximum first and second periods payoff from not declaring war; and (ii) the potential additional concession that D has to offer when T does acquire additional military capabilities. Think about the term $\frac{q^*(1-p_s)}{1-q^*p_s}$.

	Detected	Undetected
T invests	q^*p_s	$q^*(1-p_s)$
T doesn't invest	$(1-q^*)\times 0 = 0$	$(1-q^*)\times 1 = 1-q^*$

Think about the probabilities in each cell. The lower-left quadrant is zero because there is no probability that $s_1 = 1$ if T does not invest in the first place. The lower-right quadrant is $1-q^*$ because with certainty there would be no detection if T does not invest. If we add up the probabilities of the four cells, we obtain, as expected,

$$q^*p_s + q^*(1-p_s) + 0 + (1-q^*) = q^*p_s + q^* - q^*p_s + 1 - q^* = 1$$

Why do we have the term $1-q^*p_s$ in the denominator? Because we do not want to consider the case where T invests and this is detected. In such a case D would always declare war[4] and thus we should normalize by subtracting this term.

[3] Remember that the second period always results in a peaceful settlement as stated in Proposition 1.

[4] Remember that the second indifference equation RHS is about the expected payoff when D does not declare war