Collusion Among Adversaries

Matt Malis

PSSI Annual Meeting
University of Central Florida
14 November 2025

Slides: `mattmalis.github.io/slides`

The Puzzle of Provocation

Intuitive logic of provocation:

- A wants conflict, but wants B to make the first move
- A says or does something (essentially costless) to "provoke" B
- B attacks, fighting ensues

The Puzzle of Provocation

Intuitive logic of provocation:

- A wants conflict, but wants B to make the first move
- A says or does something (essentially costless) to "provoke" B
- B attacks, fighting ensues

Puzzle:

- How can it be both in A's interest to *provoke* B...
  - ...and in B's interest to *be provoked*?

The Puzzle of Provocation

Intuitive logic of provocation:

- A wants conflict, but wants B to make the first move
- A says or does something (essentially costless) to "provoke" B
- B attacks, fighting ensues

Puzzle:

- How can it be both in A's interest to *provoke* B...
  - ...and in B's interest to *be provoked*?
- How can costless communication between adversaries influence conflict behavior?

Overview

This model:

- A wants support from a third party (domestic audience, ally, etc) for conflict against B
  - needs to convince them that B has hostile intent

Overview

This model:

- A wants support from a third party (domestic audience, ally, etc) for conflict against B
  - needs to convince them that B has hostile intent
- Through cheap-talk provocation, A induces B to take an action which:
  - improves B's conflict payoffs
  - makes B more likely to fight
  - makes the third party support A's choice to fight

Overview

This model:

- A wants support from a third party (domestic audience, ally, etc) for conflict against B
  - needs to convince them that B has hostile intent
- Through cheap-talk provocation, A induces B to take an action which:
  - improves B's conflict payoffs
  - makes B more likely to fight
  - makes the third party support A's choice to fight

$\Longrightarrow$ the adversary leaders collude to manipulate the third party's behavior

Overview

Outline:

- Model setup
- Benchmark: no communication
- Private communication
- Public communication
- Cases

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Sequence:

- types θ_L, θ_F realized, obs. privately

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Sequence:

- types θ_L, θ_F realized, obs. privately
- L: private message to F, provocative ($s = 1$) or reassuring ($s = 0$)

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Sequence:

- types θ_L, θ_F realized, obs. privately
- L: private message to F, provocative ($s = 1$) or reassuring ($s = 0$)
- F: mobilize ($r_F = 1$) or not ($r_F = 0$)

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Sequence:

- types θ_L, θ_F realized, obs. privately
- L: private message to F, provocative ($s = 1$) or reassuring ($s = 0$)
- F: mobilize ($r_F = 1$) or not ($r_F = 0$)
- D: mobilize in support of L ($r_H = 1$) or not ($r_H = 0$)

Model Setup

Players L (she), D (they), F (he):

- leader L and domestic actor D within Home country H
- foreign leader F (unitary actor)

Sequence:

- types θ_L, θ_F realized, obs. privately
- L: private message to F, provocative ($s = 1$) or reassuring ($s = 0$)
- F: mobilize ($r_F = 1$) or not ($r_F = 0$)
- D: mobilize in support of L ($r_H = 1$) or not ($r_H = 0$)
- L and F (simultaneously): take aggressive action ($a_i = 1$) or not ($a_i = 0$)

Model Setup

Conflict overview:

- Conflict with strategic complements
  - Baliga & Sjöstrom (2012); Kydd (1997, 2000, 2005); Schultz (2005)

Model Setup

Conflict overview:

- Conflict with strategic complements
  - Baliga & Sjöstrom (2012); Kydd (1997, 2000, 2005); Schultz (2005)
- L and F each uncertain whether the other has SH preferences (θ_i low) or PD preferences (θ_i high)

Model Setup

Conflict overview:

- Conflict with strategic complements
  - Baliga & Sjöstrom (2012); Kydd (1997, 2000, 2005); Schultz (2005)
- L and F each uncertain whether the other has SH preferences (θ_i low) or PD preferences (θ_i high)
- D has SH preferences
  - only wants to mobilize if conflict is inevitable
  - doesn't want to increase risk of conflict by mobilizing

Model Setup

Conflict overview:

- Conflict with strategic complements
  - Baliga & Sjöstrom (2012); Kydd (1997, 2000, 2005); Schultz (2005)
- L and F each uncertain whether the other has SH preferences (θ_i low) or PD preferences (θ_i high)
- D has SH preferences
  - only wants to mobilize if conflict is inevitable
  - doesn't want to increase risk of conflict by mobilizing
- Each side wants the other side to not attack

	$a_F = 0$	$a_F = 1$
$a_H = 0$	0, 0	
$a_H = 1$		

	$a_F = 0$	$a_F = 1$
$a_H = 0$	$0,\ 0$	$-\alpha + w_i(\theta, r),$
$a_H = 1$	$\alpha + w_i(\theta, r),$	$w_i(\theta, r),$

- $w_D(\theta, r) = \theta_D + r_H \delta_H \phi$, for $\phi > 0$
- $w_L(\theta, r) = \theta_L + r_H \delta_H$

- private types $\theta_L,\ \theta_F$, with $\theta_i \sim G_i(\cdot)$ on $\left[\underline{\theta}_i, \overline{\theta}_i\right]$:
  - each leader's willingness to take the aggressive action (or, dissatisfaction with SQ)
  - θ_D common knowledge, $\theta_D < -\phi\delta_H - \alpha$

	$a_F = 0$	$a_F = 1$
$a_H = 0$	$0,\ 0$	$-\alpha + w_i(\theta, r),$
$a_H = 1$	$\alpha + w_i(\theta, r),$	$w_i(\theta, r),$

- $w_D(\theta, r) = \theta_D + r_H \delta_H \phi$, for $\phi > 0$
- $w_L(\theta, r) = \theta_L + r_H \delta_H$

- private types $\theta_L,\ \theta_F$, with $\theta_i \sim G_i(\cdot)$ on $\left[\underline{\theta}_i, \overline{\theta}_i\right]$:
  - each leader's willingness to take the aggressive action (or, dissatisfaction with SQ)
  - θ_D common knowledge, $\theta_D < -\phi\delta_H - \alpha$
- r_H, r_F: mobilization/preparation for conflict

	$a_F = 0$	$a_F = 1$
$a_H = 0$	$0,\ 0$	$-\alpha + w_i(\theta, r),$ $\alpha + w_F(\theta, r)$
$a_H = 1$	$\alpha + w_i(\theta, r),$ $-\alpha + w_F(\theta, r)$	$w_i(\theta, r),$ $w_F(\theta, r)$

- $w_D(\theta, r) = \theta_D + r_H \delta_H \phi$, for $\phi > 0$
- $w_L(\theta, r) = \theta_L + r_H \delta_H$
- $w_F(\theta, r) = \theta_F + r_F \delta_F$

- private types $\theta_L,\ \theta_F$, with $\theta_i \sim G_i(\cdot)$ on $\left[\underline{\theta}_i, \overline{\theta}_i\right]$:
  - each leader's willingness to take the aggressive action (or, dissatisfaction with SQ)
  - θ_D common knowledge, $\theta_D < -\phi\delta_H - \alpha$
- r_H, r_F: mobilization/preparation for conflict
- δ_H, δ_F: benefit from mobilizing (reduction in conflict costs)

	$a_F = 0$	$a_F = 1$
$a_H = 0$	$0,\ 0$	$-\alpha + w_i(\theta, r),$ $\alpha + w_F(\theta, r)$
$a_H = 1$	$\alpha + w_i(\theta, r),$ $-\alpha + w_F(\theta, r)$	$w_i(\theta, r),$ $w_F(\theta, r)$

- $w_D(\theta, r) = \theta_D + r_H \delta_H \phi$, for $\phi > 0$
- $w_L(\theta, r) = \theta_L + r_H \delta_H$
- $w_F(\theta, r) = \theta_F + r_F \delta_F$

- private types $\theta_L,\ \theta_F$, with $\theta_i \sim G_i(\cdot)$ on $\left[\underline{\theta}_i, \overline{\theta}_i\right]$:
  - each leader's willingness to take the aggressive action (or, dissatisfaction with SQ)
  - θ_D common knowledge, $\theta_D < -\phi\delta_H - \alpha$
- r_H, r_F: mobilization/preparation for conflict
- δ_H, δ_F: benefit from mobilizing (reduction in conflict costs)
- α: benefit of taking advantage of opponent (/cost of being taken advantage of)

	$a_F = 0$	$a_F = 1$
$a_H = 0$	$0,\ 0$	$-\alpha + w_i(\theta, r),$ $\alpha + w_F(\theta, r)$
$a_H = 1$	$\alpha + w_i(\theta, r),$ $-\alpha + w_F(\theta, r)$	$w_i(\theta, r),$ $w_F(\theta, r)$

- $w_D(\theta, r) = \theta_D + r_H \delta_H \phi$, for $\phi > 0$
- $w_L(\theta, r) = \theta_L + r_H \delta_H$
- $w_F(\theta, r) = \theta_F + r_F \delta_F$

- private types $\theta_L,\ \theta_F$, with $\theta_i \sim G_i(\cdot)$ on $\left[\underline{\theta}_i, \overline{\theta}_i\right]$:
  - each leader's willingness to take the aggressive action (or, dissatisfaction with SQ)
  - θ_D common knowledge, $\theta_D < -\phi\delta_H - \alpha$
- r_H, r_F: mobilization/preparation for conflict
- δ_H, δ_F: benefit from mobilizing (reduction in conflict costs)
- α: benefit of taking advantage of opponent (/cost of being taken advantage of)
- Assume $\delta_i > \alpha > 0$:
  - better to prepare in advance than catch your opponent off-guard

Interpretations of L/D relationship

L (agent)	D (principal)
Leader	Voter / domestic constituency

Interpretations of L/D relationship

L (agent)	D (principal)
Leader	Voter / domestic constituency
Executive	Legislature

Interpretations of L/D relationship

L (agent)	D (principal)
Leader	Voter / domestic constituency
Executive	Legislature
Protegé state	Patron state

Interpretations of L/D relationship

L (agent)	D (principal)
Leader	Voter / domestic constituency
Executive	Legislature
Protegé state	Patron state
State	International community

Interpretations of L/D relationship

L (agent)	D (principal)
Leader	Voter / domestic constituency
Executive	Legislature
Protegé state	Patron state
State	International community
Military leadership	Civilian government

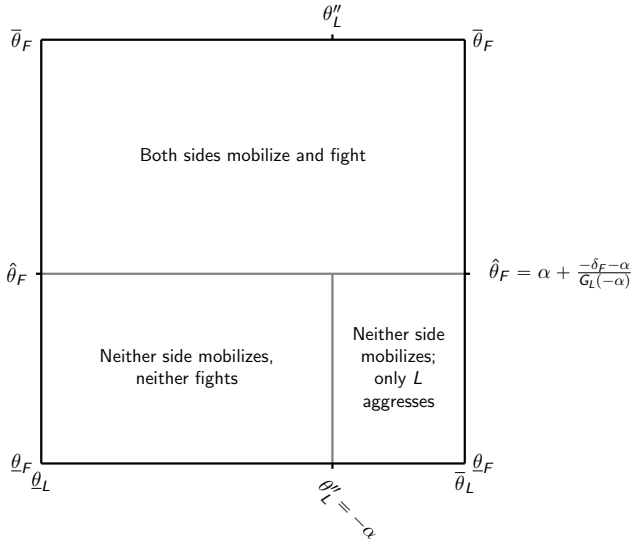
Plan for analysis

At the conflict stage, the (fight, fight) eqm, $a = (1, 1)$, is always supported

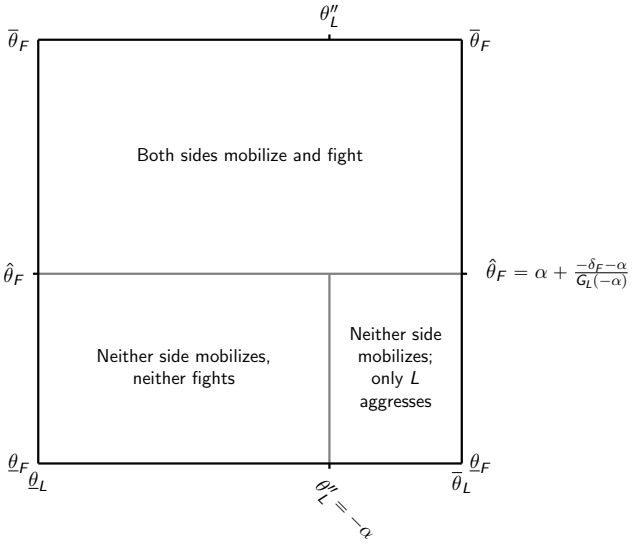
- if players anticipate $a = (1, 1)$, then both want to prepare, $r = (1, 1)$

Goal: characterize the most cooperative eqm that can be supported

Non-Communication Equilibrium

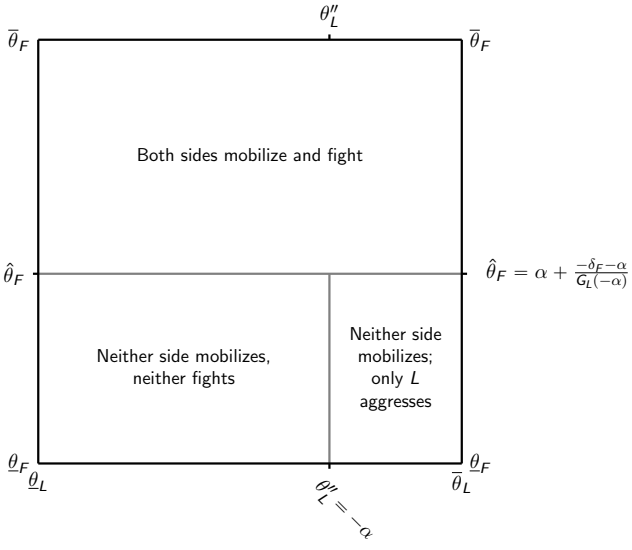


Non-Communication Equilibrium



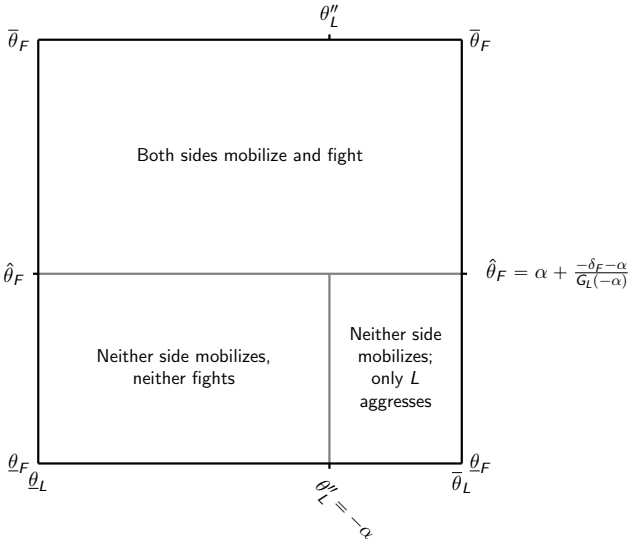
- F uncertain re:
  - will L reciprocate cooperation

Non-Communication Equilibrium



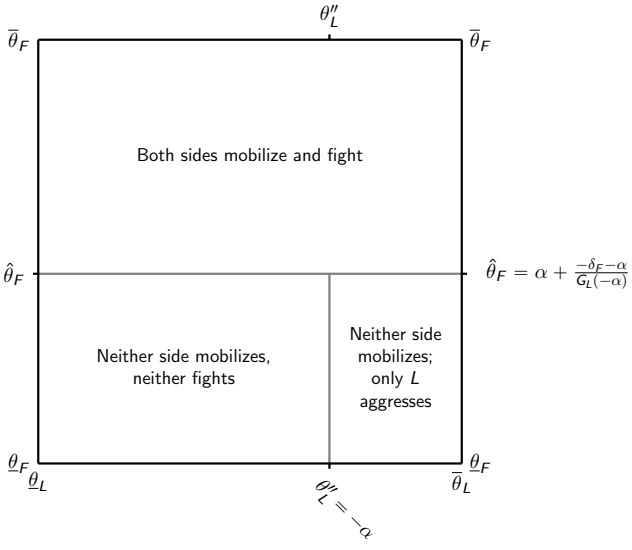
- F uncertain re:
  - will L reciprocate cooperation
- F: $r_F = 1 \iff a_F = 1$

Non-Communication Equilibrium



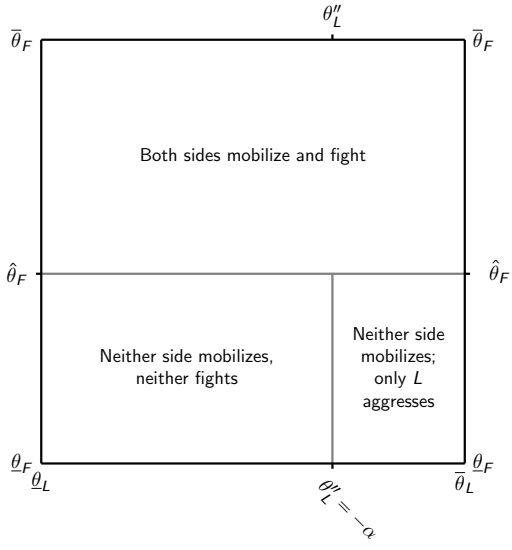
- F uncertain re:
  - will L reciprocate cooperation
- F: $r_F = 1 \iff a_F = 1$
- D: $r_H = r_F$

Non-Communication Equilibrium

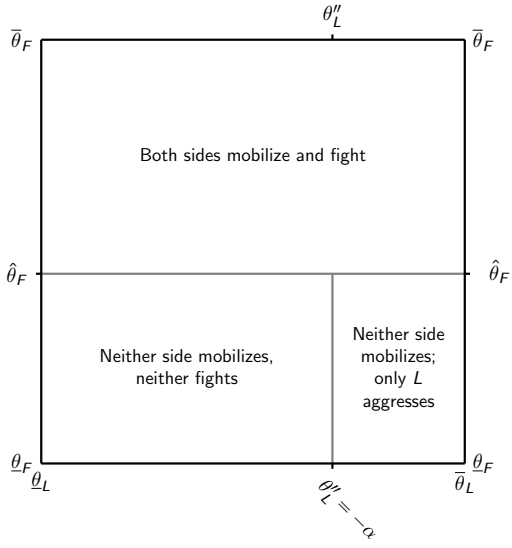


- F uncertain re:
  - will L reciprocate cooperation
- F: $r_F = 1 \iff a_F = 1$
- D: $r_H = r_F$
- L: $a_H = 0$ only if:
  - $r_F = 0$ and $\theta_L < \theta''_L$

Non-Communication Equilibrium

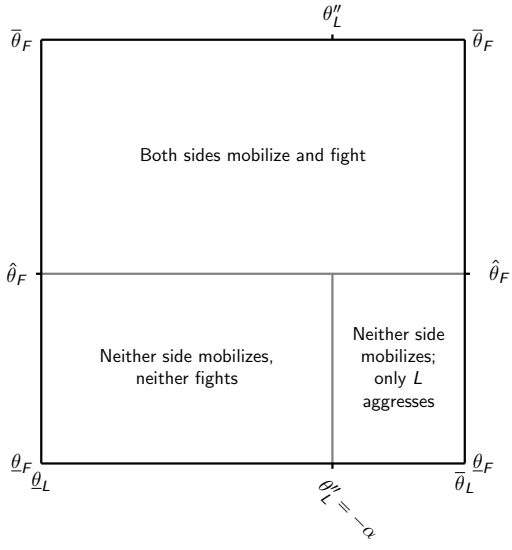


Non-Communication Equilibrium



Problems with the non-communication eqm:

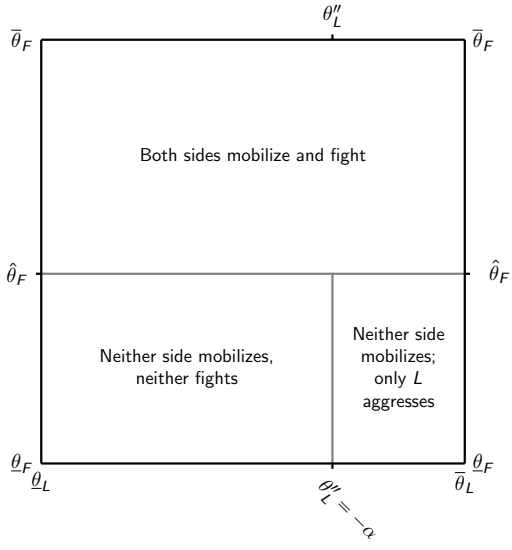
Non-Communication Equilibrium



Problems with the non-communication eqm:

- with low θ_L, moderate θ_F:
  - F mobilizes, leading to conflict
  - but everyone would prefer peace

Non-Communication Equilibrium



Problems with the non-communication eqm:

- with low θ_L, moderate θ_F:
  - F mobilizes, leading to conflict
  - but everyone would prefer peace
- with high θ_L, low θ_F:
  - conflict is inevitable
  - but F doesn't prepare
  - and L doesn't get D's support

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports
- L_h, or $\theta_L > \theta''_L$: provoke; and fight unconditionally

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports
- L_h, or $\theta_L > \theta''_L$: provoke; and fight unconditionally

- F_ℓ, or $\theta_F < \theta'_F$: never mobilize; and fight only if D mobilized

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports
- L_h, or $\theta_L > \theta''_L$: provoke; and fight unconditionally

- F_ℓ, or $\theta_F < \theta'_F$: never mobilize; and fight only if D mobilized
- F_m, or $\theta_F \in (\theta'_F, \theta''_F)$: mobilize and fight iff provoked

Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports
- L_h, or $\theta_L > \theta''_L$: provoke; and fight unconditionally

- F_ℓ, or $\theta_F < \theta'_F$: never mobilize; and fight only if D mobilized
- F_m, or $\theta_F \in (\theta'_F, \theta''_F)$: mobilize and fight iff provoked
- F_h, or $\theta_F > \theta''_F$: always mobilize and fight

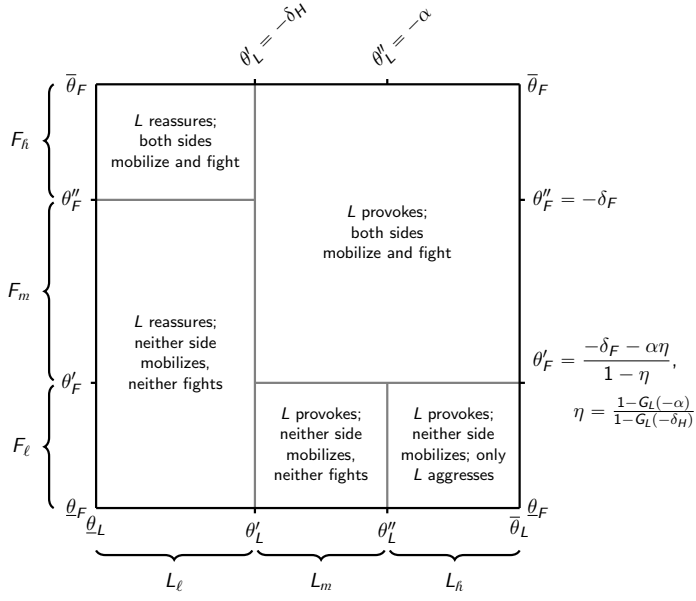
Private Communication Equilibrium

L and F will play cutpoint strategies in θ_i

- L_ℓ, or $\theta_L < \theta'_L$: reassure; and fight only if F mobilizes
- L_m, or $\theta_L \in (\theta'_L, \theta''_L)$: provoke; but fight only if D supports
- L_h, or $\theta_L > \theta''_L$: provoke; and fight unconditionally

- F_ℓ, or $\theta_F < \theta'_F$: never mobilize; and fight only if D mobilized
- F_m, or $\theta_F \in (\theta'_F, \theta''_F)$: mobilize and fight iff provoked
- F_h, or $\theta_F > \theta''_F$: always mobilize and fight

- D matches F's action


$\theta'_L = -\delta_H$
$\theta''_L = -\alpha$
$\overline{\theta}_F$
$\overline{\theta}_F$
F_h
L reassures; both sides mobilize and fight
θ''_F
$\theta''_F = -\delta_F$
L provokes; both sides mobilize and fight
F_m
L reassures; neither side mobilizes, neither fights
θ'_F
$\theta'_F = \frac{-\delta_F - \alpha\eta}{1-\eta}$,
$\eta = \frac{1-G_L(-\alpha)}{1-G_L(-\delta_H)}$
L provokes; neither side mobilizes, neither fights
L provokes; neither side mobilizes; only L aggresses
F_ℓ
$\underline{\theta}_F$
$\underline{\theta}_F$
$\underline{\theta}_L$
θ'_L
θ''_L
$\overline{\theta}_L$
L_ℓ
L_m
L_h

Private Communication Equilibrium

Mechanism:

- L_m and L_{fi} want D's support for conflict with F
  - need to convince D that F is likely to attack, even if D doesn't mobilize

Private Communication Equilibrium

Mechanism:

- L_m and L_{fi} want D's support for conflict with F
  - need to convince D that F is likely to attack, even if D doesn't mobilize
- when F gets the provocative message:
  - unsure whether L is "bluffing", or L would actually attack without F mobilizing first

Private Communication Equilibrium

Mechanism:

- L_m and $L_{\not m}$ want D's support for conflict with F
  - need to convince D that F is likely to attack, even if D doesn't mobilize
- when F gets the provocative message:
  - unsure whether L is "bluffing", or L would actually attack without F mobilizing first
- F_m mobilizes iff provoked
  - even though he knows that doing so will sway D to support

Private Communication Equilibrium

Mechanism:

- L_m and $L_{\not m}$ want D's support for conflict with F
  - need to convince D that F is likely to attack, even if D doesn't mobilize
- when F gets the provocative message:
  - unsure whether L is "bluffing", or L would actually attack without F mobilizing first
- F_m mobilizes iff provoked
  - even though he knows that doing so will sway D to support
- D supports conflict iff F mobilizes
  - even though they know F may just be reacting to L's provocation

Private Communication Equilibrium

Summary:

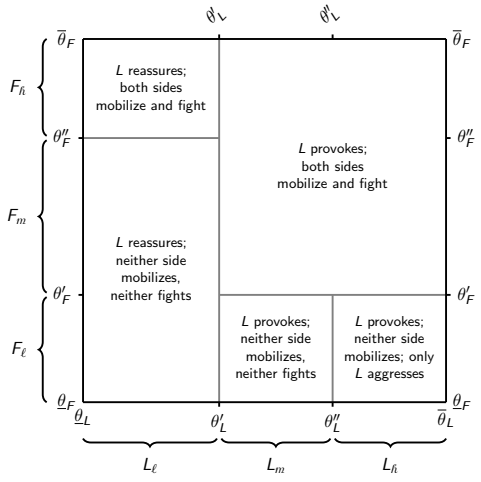
- L can enlist F's help in getting D's support for conflict
  - L and F "collude" to manipulate D's behavior

Private Communication Equilibrium

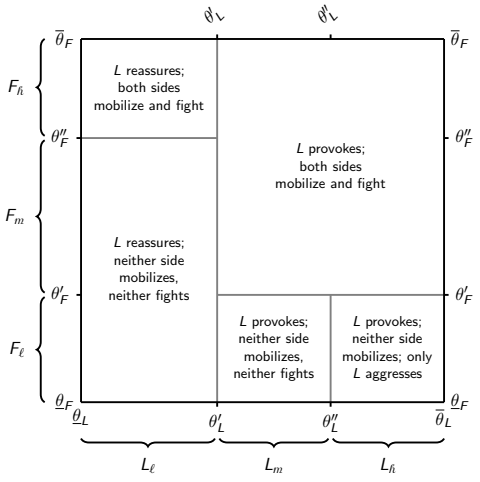
Summary:

- L can enlist F's help in getting D's support for conflict
  - L and F "collude" to manipulate D's behavior
- Corollary: if L can *provoke* F, then L can also *reassure*
  - cheap-talk message can prevent F from attacking

Comparative Statics

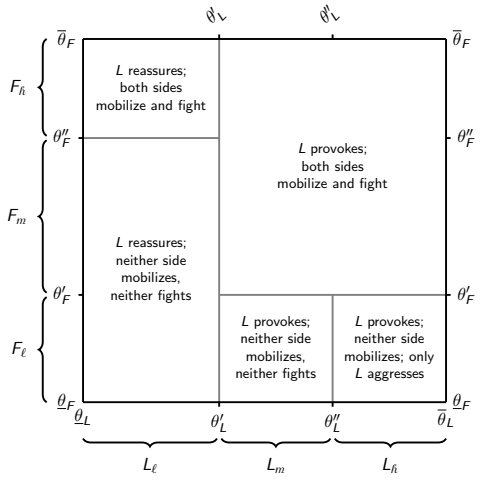


Comparative Statics



Consider increasing δ_H:

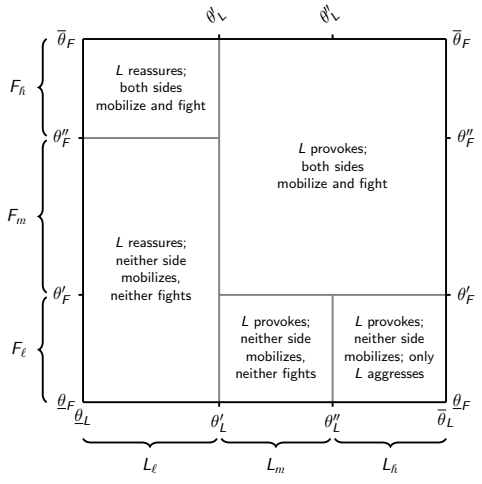
Comparative Statics



Consider increasing δ_H:

- meaning L more dependent on D's support
  - more democratic, or more concerned with international legitimacy

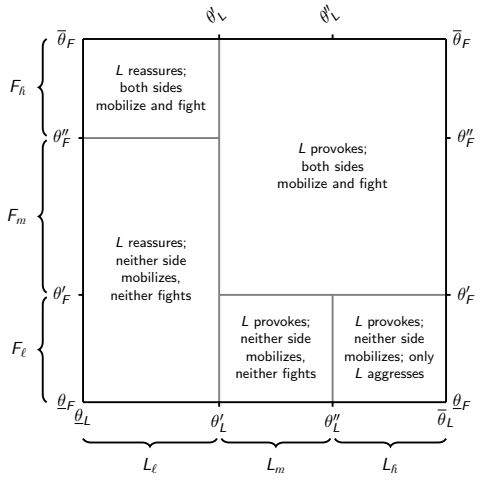
Comparative Statics



Consider increasing δ_H:

- meaning L more dependent on D's support
  - more democratic, or more concerned with international legitimacy
- Consequence of increasing δ_H:
  - L more likely to try to provoke ($\theta'_L \downarrow$)
  - L's provocation is *less* effective (i.e. F is less responsive to L's message) ($\theta'_F \uparrow$)

Comparative Statics

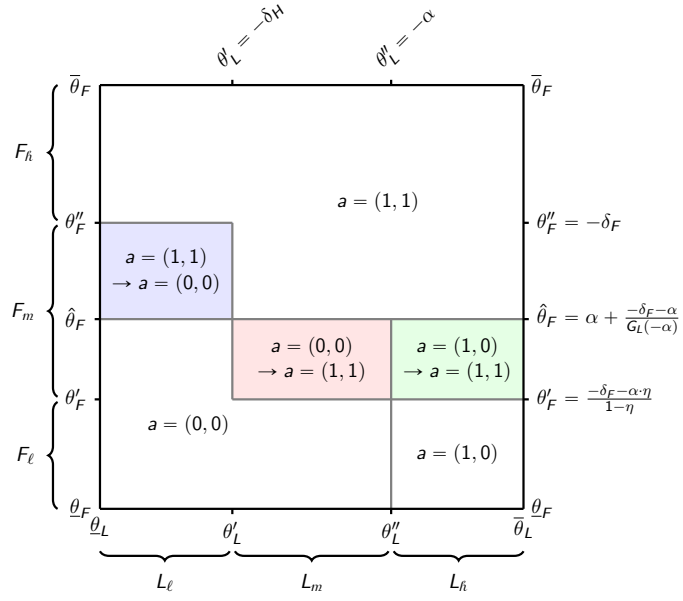


Consider increasing δ_H:

- meaning L more dependent on D's support
  - more democratic, or more concerned with international legitimacy
- Consequence of increasing δ_H:
  - L more likely to try to provoke ($\theta'_L \downarrow$)
  - L's provocation is *less* effective (i.e. F is less responsive to L's message) ($\theta'_F \uparrow$)

$\Longrightarrow$ contrary to standard audience cost result

Comparing Private Communication vs. Non-Communication



Public Communication

Alternative setup:

- L sends message publicly, D and F hear

Public Communication

Alternative setup:

- L sends message publicly, D and F hear

Result:

- Either the same behavior (if $\phi < \hat{\phi}$)
- or the message is strictly *more* effective
  - (creates more separation in F's mobilization strategy, and thus conflict
- reason: provocative message increases D' belief that conflict is inevitable

Interpretations of L/D relationship

L	D
Leader	Voter / domestic constituency
Executive	Legislature
Protegé state	Patron state
State	International community
Military leadership	Civilian government

Cases

Franco-Prussian War, 1870:

- Bismarck (L) wanted smaller German states (D) to unify under Prussia ($r_H = 1$)
  - needed to demonstrate that French Emperor Napoleon III (F) had hostile intent

Cases

Franco-Prussian War, 1870:

- Bismarck (L) wanted smaller German states (D) to unify under Prussia ($r_H = 1$)
  - needed to demonstrate that French Emperor Napoleon III (F) had hostile intent
- Ems Dispatch ($s = 1$):
  - "effect of a red rag upon the Gallic bull"
  - costless message that communicated Bismarck's hostile intent

Cases

Franco-Prussian War, 1870:

- Bismarck (L) wanted smaller German states (D) to unify under Prussia ($r_H = 1$)
  - needed to demonstrate that French Emperor Napoleon III (F) had hostile intent
- Ems Dispatch ($s = 1$):
  - "effect of a red rag upon the Gallic bull"
  - costless message that communicated Bismarck's hostile intent
- France "mobilizing" ($r_F = 1$):
  - initiating war on its own timeline, rather than waiting
  - advantageous for France — power shifting towards Prussia

Cases

Gulf of Tonkin incident, Jul-Aug 1964:

- U.S. taking provocative actions toward North Vietnam
  - ineffective covert activities (34A), commando raids, subversion attempts
  - Navy destroyer, Maddox, on radar harassment patrols
  - North Vietnamese boats attack Maddox ($r_F = 1$)
  - Johnson then sends Maddox and another ship to be attacked again

Cases

Gulf of Tonkin incident, Jul-Aug 1964:

- U.S. taking provocative actions toward North Vietnam
  - ineffective covert activities (34A), commando raids, subversion attempts
  - Navy destroyer, Maddox, on radar harassment patrols
  - North Vietnamese boats attack Maddox ($r_F = 1$)
  - Johnson then sends Maddox and another ship to be attacked again
- Congress passes Gulf of Tonkin Resolution ($r_H = 1$):
  - authorizing the President to "take all necessary measures to repel any armed attack against the forces of the United States to and to prevent further aggression"

Cases

Gulf of Tonkin incident, Jul-Aug 1964:

- U.S. taking provocative actions toward North Vietnam
  - ineffective covert activities (34A), commando raids, subversion attempts
  - Navy destroyer, Maddox, on radar harassment patrols
  - North Vietnamese boats attack Maddox ($r_F = 1$)
  - Johnson then sends Maddox and another ship to be attacked again
- Congress passes Gulf of Tonkin Resolution ($r_H = 1$):
  - authorizing the President to "take all necessary measures to repel any armed attack against the forces of the United States to and to prevent further aggression"
- Resolution substantially reduced Johnson's domestic political cost of future military escalation (δ)

Cases

South Korean martial law episode, Dec. 2024:

- Yoon government (L) wanted to overcome domestic gridlock
  - needed to demonstrate to public that NK (F) was a threat

Cases

South Korean martial law episode, Dec. 2024:

- Yoon government (L) wanted to overcome domestic gridlock
  - needed to demonstrate to public that NK (F) was a threat
- "Provoke the North's attack at the NLL" ($s = 1$)
  - drone flights dropping propaganda leaflets
  - shooting down trash balloons

Cases

South Korean martial law episode, Dec. 2024:

- Yoon government (L) wanted to overcome domestic gridlock
  - needed to demonstrate to public that NK (F) was a threat
- "Provoke the North's attack at the NLL" ($s = 1$)
  - drone flights dropping propaganda leaflets
  - shooting down trash balloons
- Provocation unsuccessful; Kim Jong Un did not mobilize
  - Yoon's martial law attempt was overwhelmingly rejected by SK public ($r_H = 0$)
    - did not perceive external threat

Contribution

Distinction from **audience costs**:

- message is intended to *provoke*, rather than *deter* as in AC
- message can be private or public
- strategy is *less* effective for democratic leaders

Contribution

Distinction from **audience costs**:

- message is intended to *provoke*, rather than *deter* as in AC
- message can be private or public
- strategy is *less* effective for democratic leaders

Distinction from **diversionary war**:

- insufficient for L to show hawkishness / competence in conflict
- must also show that the international environment (i.e. F's type) makes those attributes valuable

Contribution

Distinction from **audience costs**:

- message is intended to *provoke*, rather than *deter* as in AC
- message can be private or public
- strategy is *less* effective for democratic leaders

Distinction from **diversionary war**:

- insufficient for L to show hawkishness / competence in conflict
- must also show that the international environment (i.e. F's type) makes those attributes valuable

Novel mechanism of cheap-talk diplomacy between adversaries:

- coordinating action to collude against a third party