SUPPLEMENTARY MATERIALS

Four Costly Signaling Mechanisms

Appendix 1

- Experimental Instructions

Appendix 2

- Sample Demographics and Balance on Pre-Treatment Covariates

Appendix 3

- How the Game Changes Across Different Signaling Mechanisms

APPENDIX 1

Experimental Instructions

The experiment was administered over the Internet to a national sample of U.S. adults (aged 18 and above). The experimental design and implementation are described in the main text. The experimental instructions and manipulations are presented below screen by screen.[1]

[Screen 1: All Respondents]

In international politics, it is sometimes difficult to know the true intentions of other countries.

[Screen 2: All Respondents]

Countries can send signals to communicate their intention.

"Signals" can be words or actions.

They can be costless or costly for the country sending the signal.

[Screen 3: All Respondents]

In the next screen, you will read about a situation of international conflict that had occurred in the past and might occur again in the future. We will describe the situation and ask for your opinion.

[1] Voluntary and informed consent was obtained via the participation consent form at the beginning of the survey. Participants were provided information about payment and the duration of the survey, so that they could make an informed choice before giving their consent to participate. Participants were compensated through Amazon.com following standard practices based on previous studies. Recent experiments that also used AMT include Carlson (2019, *APSR*); Coppock, Ekins and Kirby (2018, *QJPS*); McConnell, Margalit, Malhotra and Levendusky (2018, *AJPS*); Mummolo and Peterson (2019, *APSR*). To protect anonymity, participants were not asked any question that would reveal their personal identities. No deception is used in this study.

[Screen 4: All Respondents]

Two countries – Country X and Country Z – have a dispute over a piece of territory.

Country X sends a Signal to Country Z threatening to go to war if Z does not withdraw from the territory.

[Screen 5: Randomized Conditions]

Consider a 0-10 scale from costless (0) to extremely costly (10).

[*For respondents randomized into the Sunk-Cost condition*]

In sending the signal ...

- X has paid a cost of [0 / 2 / 10]. The cost cannot be recovered regardless of whether X fulfills its threat.

[*For respondents randomized into the Tied-Hands condition*]

In sending the signal ...

- X will pay a cost of 0 if X fulfills its threat, and a cost of [0 / 2 / 10] if X does NOT fulfill its threat.

[*For respondents randomized into the Installment-Cost condition*]

In sending the signal ...

- X will pay a total cost of [0 / 2 / 10] in installments over time. Once the cost is paid, it cannot be recovered regardless of whether X fulfills its threat.

[*For respondents randomized into the Reducible-Cost condition*]

In sending the signal ...

- X has paid a cost of [0 / 2 / 10]. X gets back a value of [0 / 2 / 10] if X fulfills its threat, and 0 if X does NOT fulfill its threat.

[All Respondents]

If Z does not withdraw, do you think X is likely or unlikely to fulfill its threat?

o Likely
o Unlikely
o Neither likely nor unlikely

[Screen 6: All Respondents]

[*If "Likely"*] Do you think that it is very likely, or only somewhat likely, that X will fulfill its threat?

o Very likely
o Somewhat likely

[*If "Unlikely"*] Do you think that it is very unlikely, or only somewhat unlikely, that X will fulfill its threat?

o Very unlikely
o Somewhat unlikely

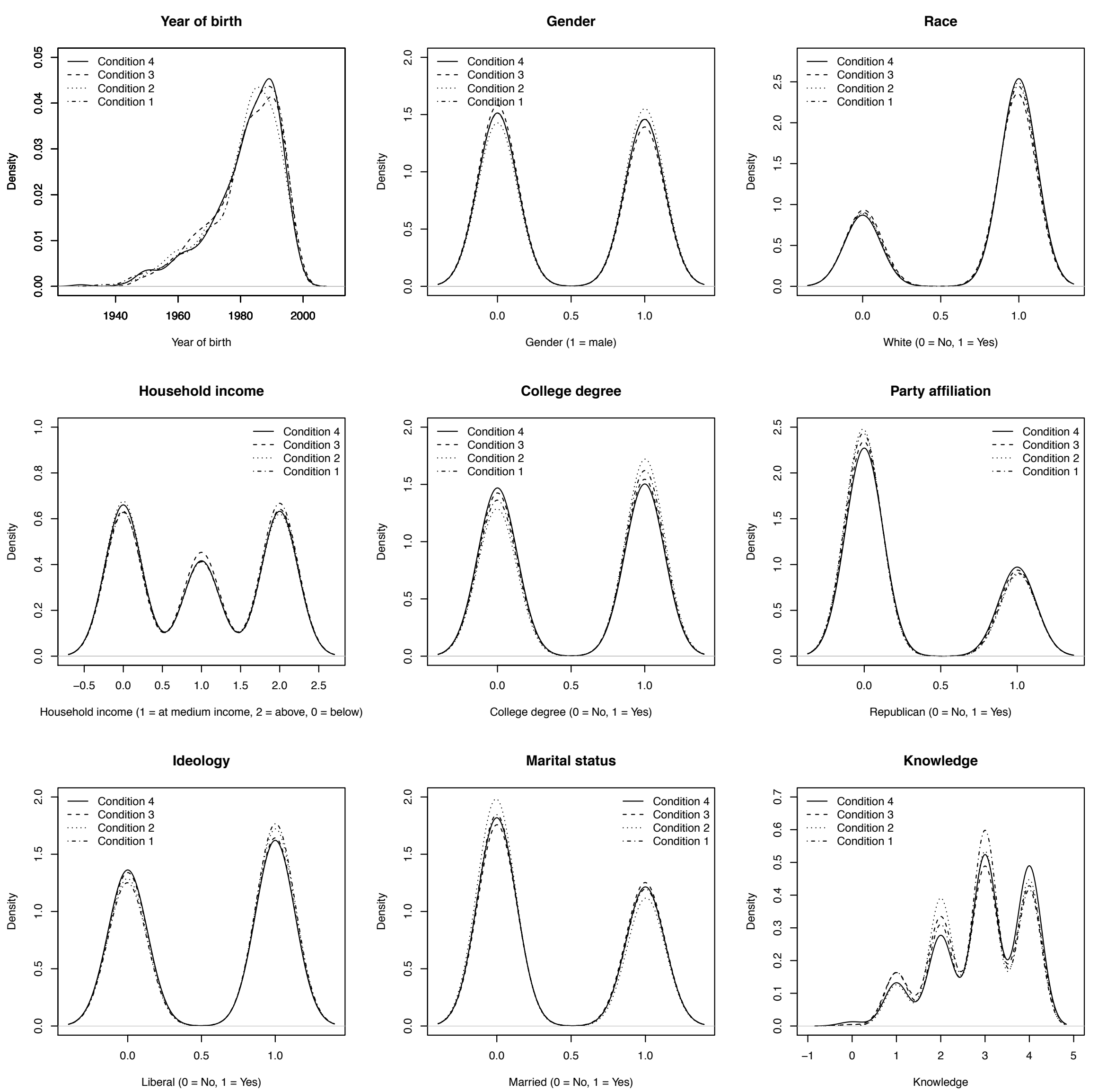
[*If "Neither likely nor unlikely"*] Do you lean toward thinking it is likely that X will fulfill its threat, lean toward thinking it is unlikely, or don't you lean either way?

o Lean toward likely
o Lean toward unlikely
o Lean neither way

APPENDIX 2

Sample Demographics and Balance on Pre-Treatment Covariates

The graphs below show the univariate distributions of a broad range of pre-treatment covariates across the four experimental subgroups: Condition 1 (Sunk Cost), 2 (Tied Hands), 3 (Installment Cost), 4 (Reducible Cost). There is no imbalance in the demographic characteristics across the four groups.



APPENDIX 3

How the Game Changes Across Different Signaling Mechanisms

This appendix highlights one way by which game forms and payoffs can change across different signaling mechanisms. Figures 1 and 2 are based on Fearon (1997). Nature informs Defender (D) and Challenger (C) of their values for the issue in contention. v_D is D's value drawn from the cumulative distribution $F_D(v)$ on the positive reals; v_C is C's value drawn from $F_C(v)$. In each game, D chooses a signal $m \geq 0$ observed by C, and C chooses whether to challenge. If C does not challenge, the game ends. If C challenges, D will decide whether to fight. If fighting occurs, D wins with probability $p \in (0,1)$, and state i's costs for war is c_i.

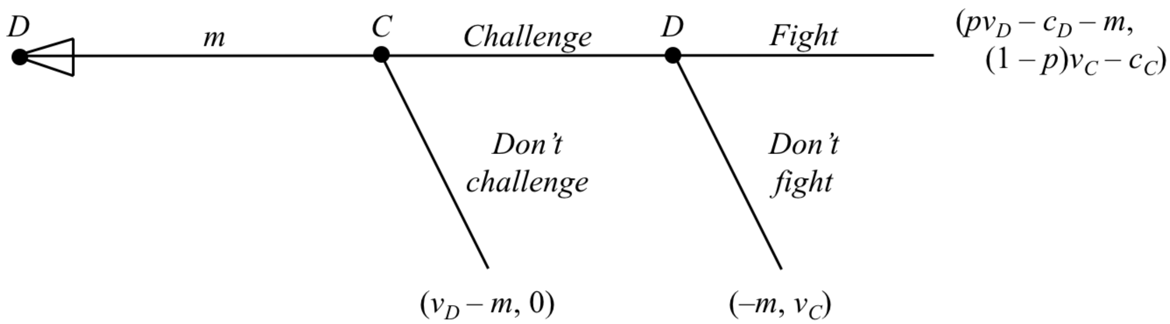


Figure 1: Sunk Cost Signals

Note: Adapted from Fearon (1997).

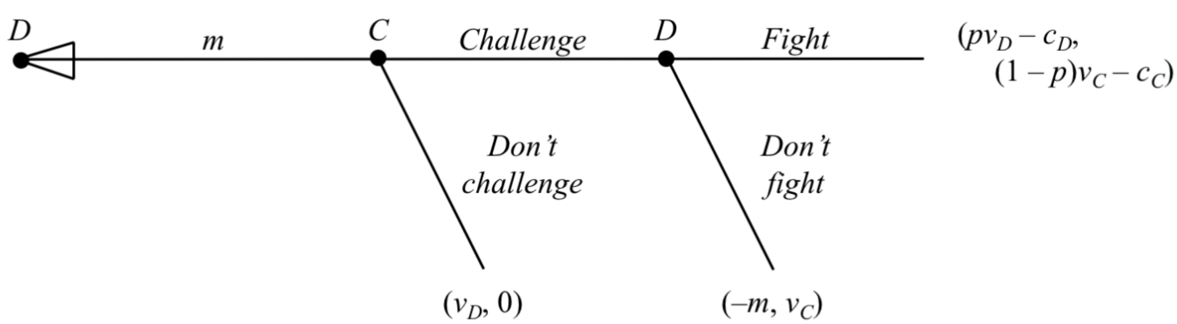


Figure 2: Tying-Hands Signals

Note: Adapted from Fearon (1997).

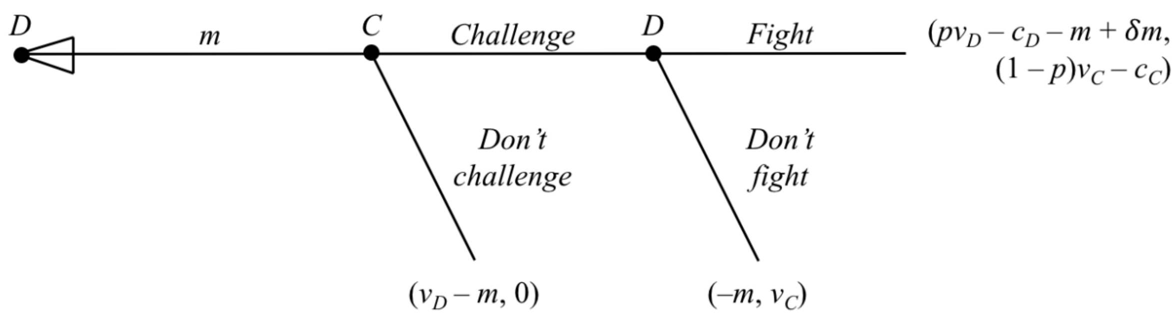


Figure 3: Reducible Cost Signals

In existing work that considers military signaling as a mixed case of sinking costs and tying hands, the signal changes the probability of victory p (Fearon 1997; Slantchev 2005). For the game in Figure 3, the reducible cost signal changes the signaler's cost of fighting but does not change p. We can think about this in various ways. First, there are investments that only affect the cost of war but not the probability of victory. Second, there are cases where military investments – whether early or late – lead to the same fighting capacity, but early investments can reduce the total costs suffered. Third, the costs are political rather than military. An example is the domestic political costs of taxing citizens to increase "guns" at the expense of "butter." Here the political cost to the leader is fixed if there is no fighting but reduced if otherwise (affirming the leader's earlier decision and reversing the political cost).

A reducible cost signal directly affects only D's war payoff but not C's, in contrast to how military signaling (e.g. mobilization) is conceptualized in the literature, where the costs of signaling affect the posterior balance of power and the war payoffs of both D and C (Fearon 1997; Slantchev 2005). Also, unlike tying hands (Fearon 1997), the cost of fighting can be shaped by the signal cost (m). Thus, the risk of war and the equilibrium level of m can change as a consequence.

The equilibrium level of m depends on the delta parameter (in δm) in the war payoff. In the case where $\delta = 1$ (full recoverability), signaling is costless if C challenges and D fights, but costly if C does not challenge. By contrast, in tying hands, signaling is costless if C does not challenge, but costly if C challenges and D does not fight. For simplicity, and because of the factors beyond D's control (e.g. geography, institutions and the state of technology), we assume δ is exogenous here. Future work may consider cases where the sender can influence the parameter in different degrees.

Installment cost signals

The installment cost signaling case is more complicated as the costs are accumulated dynamically over a time horizon, but with the installment cost = 0 at period $t = 0$. As discussed in the paper, the economics of time horizons can be modelled in many ways, and installment costs with and without commitment problems can also be formulated in different variants. The first question is what time horizon is more appropriate. This depends on the specific application. For an immediate deterrence case with a well-defined final period, a finite horizon is appropriate. For a general deterrence case with no final period salient in the actors' calculations, an infinite horizon is applicable.

In general terms, we can have a dynamic model with flow payoffs and periods $t = 0, 1, 2, \ldots$, where D is expected to incur installment cost m_t at period t where $t > 0$. If D fights at period t, the payoff is based on a discounted sum of v_D up to period t and a lottery on a future stream. Assume that D can predetermine the level of installment costs. For a pure signaling case,[2] if there is an equilibrium level of installment costs, then D may choose to incur installment costs until that equilibrium level is reached at a particular stage. In a continuous-time model, D may pay installment costs until the equilibrium stop-time, and then the installment costs stop when type separation is achieved.

This formulation assumes there is no private information about C's type. If such private information exists, then problems of "dynamic deterrence" will arise,[3] which may generate a different set of interesting implications.[4]

---

[2] There will be other interesting dynamics in a case where D is choosing signal costs (m) that have militarily relevant capabilities.

[3] Defender (D) may infer something about Challenger (C)'s type after observing C's challenge/non-challenge in a previous period, and C may infer something about D's type after observing D's installment cost in a previous period.

[4] I thank James Fearon for a discussion that clarified my thoughts. The term "dynamic deterrence" was previously used by Professor Fearon, from whom I borrowed.