*Title Page

Does Cheap Talk Matter? An Experimental Analysis

Dustin Tingley
Ph.D. Student, Politics
130 Corwin Hall, Princeton University
Princeton, NJ 08544
dtingley@princeton.edu

Barbara Walter
Professor of Political Science, UCSD
9500 Gilman Drive
La Jolla, CA 92093-0519, Office #1323
bfwalter@ucsd.edu

Abstract

What effect does cheap talk have on behavior in an entry-deterrence game? We shed light on this question by using incentivized laboratory experiments of the strategic interaction between defenders and potential entrants. Our results suggest that cheap talk can have a substantial impact on the behavior of both the target and the speaker. By sending costless threats to potential entrants, defenders are able to deter opponents in early periods of play. And after issuing threats, defenders become more eager to fight. We offer a number of different explanations for this behavior. These results bring fresh evidence about the potential importance of costless verbal communication to the field of international relations.

Does Cheap Talk Matter? An Experimental Analysis

Abstract

What effect does cheap talk have on behavior in an entry-deterrence game? We shed light on this question by using incentivized laboratory experiments of the strategic interaction between defenders and potential entrants. Our results suggest that cheap talk can have a substantial impact on the behavior of both the target and the speaker. By sending costless threats to potential entrants, defenders are able to deter opponents in early periods of play. And after issuing threats, defenders become more eager to fight. We offer a number of different explanations for this behavior. These results bring fresh evidence about the potential importance of costless verbal communication to the field of international relations.

Most bargaining models assume that verbal threats or promises that exact no costs on the sender will have little or no influence on those receiving the message.[1] A state leader can claim that she will "fight hard" when elected, or "cut taxes once in office," or "come to the aid of an ally" that is attacked, but in the absence of any punishment for not following through, these statements are generally viewed as empty and inconsequential.[2]

In reality, however, leaders engage in what could be construed as cheap talk all the time. President Kennedy promised Soviet Premier Nikita Khrushchev that he would remove American nuclear missiles from Turkey if the Soviets first removed their missiles from Cuba. France and Great Britain promised to help Poland and Czechoslovakia should the Germans attack. Secretary of State Dean Acheson claimed that the United States would not protect Korea in the 1950's. And President Clinton threatened to bomb North Korea if they continued to acquire nuclear weapons capabilities. Each of these leaders engaged in cheap talk despite the fact that there was little reason to believe that any of these pronouncements were true.

Cheap talk also appears, at times, to work. Kennedy's promise to withdraw U.S. missiles in Turkey is widely believed to have convinced Khrushchev to withdraw his

---

[1] See, for example, (Fearon 1994; Schultz 1998; Fearon 1995). For models that consider how cheap talk might be made costly through domestic institutions and/or reputational concerns, see (Ramsay 2004; Guisinger and Smith 2002; Sartori 2002). For models where cheap talk can influence behavior see (Crawford and Sobel 1982; Crawford 2003; Sobel 1985; Farrell and Gibbons 1989; Farrell 1995).

[2] The literature does, however, argue that cheap talk could be effective if there is adequate overlap in the interests of the signaler and signalee (e.g., (Kydd 2003)).

missiles from Cuba. Clinton's threat to bomb North Korea (together with promises of energy assistance) did appear to convince the North Koreans to stop their nuclear development program, at least temporarily.[3] And in their study on bargaining, Farrel and Gibbons found that "[t]alk is ubiquitous and is often listened to, even where no real penalty attaches to lying, and where claims do not directly affect payoffs" (1989, pg. 222). Verbal claims about one's intentions may be costless, but leaders frequently use them, and they appear to influence behavior in ways we do not fully understand.

This article has two goals. The first is to determine whether costless communication has any effect on behavior when used in an entry-deterrence game. If a defender is allowed to issue a verbal threat that is both costless and private, does this change the entrant's and the defender's behavior in any way? We use an incentivized laboratory experiment and find that cheap talk signals can influence behavior despite our subjects having opposing preferences. The second is to theorize about why such communication might be significant even if everyone knows it is costless. Here we consider the role honesty and common knowledge play in shaping strategic decisions.

Empirically, there are at least three ways to determine whether leaders rely on costless signaling, and if they do, whether these messages are persuasive. The first is to collect and analyze observational data. One could, for example, study all verbal communication that took place between the United States and the Soviet Union over nuclear weapons during the Cold War to see whether these messages influenced either party in any way. The problem with such a study is that it suffers from two difficult-to-resolve methodological problems. The first is that cheap talk games tend to be sensitive

[3] For more examples see Davis (2000).

to initial beliefs and controlling for these beliefs is hard to do in large N studies. Khrushchev, for example, may have already developed a reputation for toughness when he began communicating with Kennedy. The second is that cheap talk and costly signaling often co-occur, making it difficult to isolate and identify the independent effects of the very cheapest form of communication.[4] Observational data, therefore, tends to be unreliable.

A second approach would rely on qualitative case studies to trace when and how leaders engage in cheap talk and its potential effects on behavior. Studies do exist that look at relations between countries and include cheap talk as indicators, but these studies do not attempt to isolate the effect of these statements on behavior (e.g., see (Foster 2006)). Even if a study did situate itself in the bargaining literature, this approach would also have disadvantages. A small number of case studies can confirm whether individuals in those cases communicated with each other in a costless way, and if those messages had any effect, but they could not confirm whether this behavior was more widespread.

A third approach – laboratory experiments – circumvents both these problems. In a laboratory experiment the researcher can isolate costless signals and their consequences, while controlling for confounding factors. In this way, the experiment can

---

[4] One exception is Thyne (2006) which is one of the few papers that attempts to isolate cheap talk in a large N study. Schrodt (1993) presents a time series showing the dynamic of US-Soviet Relations, US-China Relations and Israeli-Palestinian relations using the COPDAB and WEIS datasets. However, cheap talk events are mixed with other events and are, therefore, difficult to evaluate.

reveal whether threats and promises are actually used, whether they directly changed behavior, and if they did change behavior, under what conditions. Laboratory experiments, however, are not without their own drawbacks. Since subjects tend to be undergraduate students as opposed to state leaders, the findings cannot be generalized to field settings. It is possible that state leaders use verbal communication differently from undergraduates even when placed in a similar context.[5] Still, an empirical test of cheap talk in the laboratory will reveal whether the hypothesized relationships emerge under ideal conditions, and help advance the debate beyond the question of whether cheap talk matters, to a more constructive discussion of when, how and why it might be used.

In what follows, we set up a simple experiment to determine whether individuals engage in cheap talk and if they do, whether it changes behavior. The experiment compares how individuals conduct themselves in an entry-deterrence game with one-sided incomplete information when cheap talk is not allowed and when it is. What we find is surprising. Verbal threats had significant effects on the behavior of both the sender and the target. Even though threats were completely costless, targets were more likely to back down if they received a threat, and senders were more likely to act on those threats. In short, when individuals engaged in cheap talk in the laboratory– and they almost always did when given the chance – it changed the behavior of everyone involved. This suggests that even the most costless form of verbal communication can be influential, at least in certain circumstances.

[5] In the absence of additional testing this possibility cannot be ruled out. Ours is a first step in that direction.

The remainder of the paper is broken down into four sections. The first section reviews current theories and findings on cheap talk in both the international relations and economics literature. Section two introduces our experimental design, presents some theoretical predictions, and explains our empirical strategy. Section three reveals the results of these experiments and offers an explanation for why cheap talk is powerful even though most bargaining models would not expect it to be. Here we highlight the potentially important roles that honest and incompetent individuals can play in changing the incentives of the game. In the final section we discuss the contributions this study makes to international relations, as well as avenues for future research.

I. What We Know Theoretically and Empirically About Cheap Talk in IR

The international relations literature has been divided between those who argue that costless verbal communication can be informative and those who argue that it provides little or no information at all. In one camp are the constructivists, who assert that things like persuasion, argumentation, and rhetoric can play a critical role in politics and diplomacy. According to Finnemore and Sikkink, "IR scholars have tended to treat speech either as "cheap talk," to be ignored, or as bargaining, to be folded into strategic interaction. However, speech can also persuade; it can change people's minds about what goals are valuable and about the roles they play (or should play) in social life" (2001, pg. 402)[6] Significant anecdotal evidence seems to support this camp's view. Throughout history, state leaders have engaged in all sorts of verbal and symbolic communications even if, on the surface, it appears shallow.

[6] See also (Risse 2000).

Formal models of interstate relations, however, consistently find that costless communication or "cheap talk" should not matter. [7] If two states have opposing preferences and incomplete information about each other's payoffs, costless messages provide no additional information about what the sender is likely to do. This is because all players have incentives to make similar claims whether they are true or not. It is only when real costs are attached to the messages that sincere senders can be distinguished from those who are just bluffing (Fearon 1995, pg. 396).

Existing empirical studies suggest that the formal models are correct. In a study of militarized disputes between 1816 and 1993, Sartori (2005) found that verbal communication in the form of diplomacy could change an adversary's mind about its desire to fight, but only if the sender had already invested heavily in the credibility of these messages through the costly use of force. Thyne (2006) found that cheap signals could actually have negative consequences. In a study of civil wars, he found that negotiations were more likely to fail if one of the disputants used hostile costless signals.[8]

The only evidence in favor of cheap talk comes from laboratory experiments where the preferences of the sender and the target are aligned. Cooper et al. (1989) and Crawford (1998) found that in a battle of the sexes game, costless communication

[7] Though some recent work in IR suggests that increasing the number of bargaining dimensions allows cheap talk to be informative when bargaining over a single dimension would not allow for this (Trager 2009).

[8] Thyne theorized that this was in part because parties were more likely to make excessive demands when costless communication was used (pg. 957).

allowed players to coordinate on an outcome, making successful cooperation possible.[9] Similarly, in a public goods game with incomplete information about private endowments, Palfrey and Rosenthal (1991) found that subjects regularly conditioned their behavior on the cheap talk message they received, but did not obtain more efficient outcomes as a result.[10] Majesky and Fricks (1995) found that cooperation was more likely in a prisoner's dilemma game if cheap talk was allowed. In all these cases, cheap talk worked, but only because each side already had an interest in cooperating.

The problem is that many interactions in the world of international relations occur under conditions where players have conflicting preferences. World leaders often benefit from deceiving and misleading each other and frequently do not want to cooperate. Not surprisingly, the few experiments structured around conflictual situations have failed to find that cheap talk had any lasting influence on behavior. Forsythe and colleagues, for example, found that in a bargaining game with one-sided incomplete information

---

[9] The form of costless communication may also matter. Isaac and Walker (1988) and Ostom et al. (1992) found that in a public goods provision game, subjects were more likely to contribute larger sums of money if verbal pledges were made face-to-face rather than anonymously.

[10] However, subjects did not ultimately obtain more efficient outcomes as a result. In other experiments, Wilson and Sell found that subjects contributed more in a public good game when pre-play communication was allowed *and* information existed about past behavior (1997). Wilson and Sell did, however, find, surprisingly, that subjects contributed the most when they could not communicate with each other and had no information about past behavior.

individuals did not behave much differently if they were allowed to communicate cheaply versus if they were not allowed to communicate at all (Forsythe et al. 1991).[11] Similarly, Croson and colleagues found that in an ultimatum game with incomplete information about outside options cheap talk affected behavior, but only temporarily.[12] Subjects could increase their short term bargaining outcomes by using cheap talk, but would be punished in the long term if they chose to lie (Croson et al. 2003). [13] Finally, in an n-person market entry game, Sundali and Seale (2004) found that entrants exaggerated their intention to enter when given the chance, but that this did not influence how others played the game. The balance of experimental results, therefore, suggests that cheap talk will have very little influence on behavior in more conflict-prone settings.

In what follows, we investigate the effects of cheap talk in an experiment that more closely models a wider range of IR interactions. Specifically, we examine the influence of cheap talk when there are incentives to build reputations that could influence

[11] Uncertainty in this game was over the size of a resource to be divided.

[12] This was the case if it was possible to detect lying.

[13] Our interest is in line with that of Croson and colleagues in that we are interested in the role of cheap talk in bargaining environments. Our investigation differs from theirs in several important ways. They looked at the role of reputation between a pair of actors who repeatedly interacted with each other. Our study looks at behavior where a single "defender" faces a series of different challengers ("strangers" design as opposed to a "partners" design). The strategic game we study also differs. They use a repeated ultimatum game with outside options, whereas we use a repeated entry-deterrence game.

the choices of *future* actors.[14] We believe that by examining the role of cheap talk in a common strategic situation in international relations, we can begin to understand the puzzle state leaders present for our theoretical models. If cheap talk really serves little positive purpose in most conflict situations, why do world leaders so frequently use it?

III. Cheap Talk and Entry-Deterrence

In what follows, we introduce a game in international relations that allows us to study the effects of cheap talk in situations where players have strong incentives to deceive each other, especially in early periods of the game. We have chosen an entry-deterrence game for three reasons. First, it is relatively common in international affairs for governments to use verbal threats as part of an attempt to deter potential challengers. China's verbal pronouncements against any move by Taiwan to declare independence, or its threats against separatist regions are real-world examples of this type of game. Second, an entry-deterrence game has a simple sequential structure which allows us to observe when defenders choose to issue threats, and how different entrants react to any threat they may receive. Finally, the experimental economics literature on reputation building is surprisingly quiet on the role of cheap talk in this type of repeated bargaining environment (see below). Thus, there are good substantive and methodological reasons for choosing this particular game.

[14] Our experiment differs from Sundali and Seale (2004)– the most closely related experiment - in several respects. First, our defenders faced a sequence of entrants. In their experiment everyone played the same role (entrant) and decided whether or not to enter a market. Second, there is no incomplete information or chance for reputation building.

We begin by presenting the simple game of one-sided incomplete information. We then characterize the sequential equilibrium of a repeated version of the game with no communication (and hence no cheap talk), and then consider how we would expect cheap talk to influence behavior.

The Structure of the Game

The game is straightforward. In it, a defender faces a series of potential entrants who must decide whether to challenge the defender or stay quiet. The defender, in turn, must decide whether to fight entry or allow the challenger to enter. Figure one reveals the structure of a single-shot play of the game as well as the payoffs each of the players knows it will receive for the different outcomes.[15]

Figure 1: The Structure of a Single-Shot Play

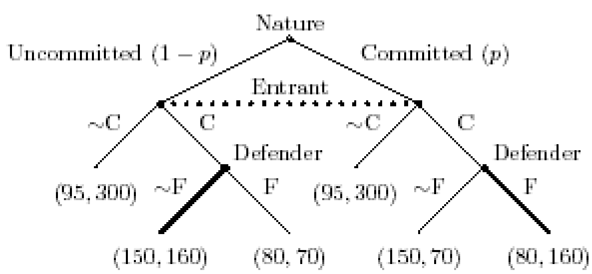


Figure 1: The Structure of a Single-Shot Play

The game begins with nature randomly choosing whether the defender is committed (strong) or uncommitted (weak) to fighting a challenge with probability p. This introduces the element of uncertainty necessary for reputation building to occur. If the defender is committed, it will always prefer to fight entry rather than acquiesce since this will always deliver better payoffs (see Figure 1). If it is uncommitted, it would prefer

[15] Payoff parameters are from Jung, Kagel and Levin (1994).

to acquiesce rather than pay the costs of war.[16] Once nature has chosen the defender's type, the entrant must decide whether to challenge (C) or remain not challenge (~C).

The key to the game is that the entrant does not know whether it is facing a defender who is committed (in which case the entrant would prefer not to challenge), or a defender who is uncommitted (in which case the entrant would prefer to challenge). If the entrant decides to challenge, the defender then chooses whether to fight (F) this challenger or concede (~F).

In the repeated play version of this game, once the defender makes his or her choice, a second entrant then chooses whether to challenge, after which the defender again decides whether to fight or accommodate. As each entrant plays, they obtain information about how previous entrants played against the defender they are currently matched with, and how the defender played *if* the previous entrant decided to challenge. Thus, they are able to update their beliefs about what type of defender they are likely to face. The game continues until the defender has been pitted against a commonly known number of entrants.[17] How the defender behaves toward an early entrant, therefore, can

[16] In this case, the payoffs are 160 for not fighting a challenger and 70 for fighting a challenger.

[17] Walter (2006) discusses cases with multiple different entrants. In order to keep the framework consistent with earlier work on the entry-deterrence game we only analyze repeated play between different opponents. The game could also be played repeatedly between a defender and a single entrant. This would be similar to a situation where a government engaged in a series of continuing disputes with a single ethnic group, where

be interpreted as important information about how the defender is likely to behave toward later entrants.

For our analysis of the role of cheap talk, we had our subjects play the game two different ways. In one version, they engage in the game exactly as we described it without any communication between the defender and entrants. In the second version, defenders are given the opportunity to issue a costless threat. Our test of cheap talk, therefore, entailed a simple addition to the game. Each defender sent a signal to each potential entrant. They could either issue a message that said they would fight if faced with entry or they could send a message that said that they would not fight. [18]

A critical feature of this communication is that it is private. No other player other than the current challenger was able to see the message. This allowed us to observe the

---

the ethnic group demanded greater and greater concessions over time (e.g., the Canadian government's relationship with the Parti Quebecois).

[18] We considered a number of other options for the message space. We could have allowed subjects to select not sending a message at all. Or we could have allowed the subjects to choose a costly signal. While each of these additions would have improved the correspondence between the communication options available to real decision makers and those in our experiment, each would have introduced their own complications both theoretically (cheap talk models typically assume that *some* message is sent) and empirically (conditioning our analysis on three or four message options instead of two would exacerbate sample size problems). Others in this experimental literature share our approach (Palfrey and Rosenthal 1991, pg. 188). For an interesting study of open communication and threats in an IR simulation experiment, see McDermott et al. (2002).

messages in their most costless form. Since none of the messages can be observed by any other player – a fact made very clear both in our instructional period and during the experiment—there were no incentives for the sender to follow through with threats for reputational reasons.[19] Thus, senders gained no additional deterrent value by publicly validating their threats. This creates a situation where talk is truly "cheap" and no costs can be imposed on the sender for not following through.

Theoretical Predictions About Cheap Talk

If no cheap talk is allowed, the standard entry-deterrence model – as outlined by Kreps and Wilson (1982) and Milgrom and Roberts (1982) –makes three predictions about how the defender should play.[20] First, strong-type defenders should always fight no matter what period they are in. Second, weak defenders should play a strategy that depends on how many entrants remain. Weak defenders know that if they acquiesce to the first challenger this will immediately reveal their type and this information will trigger a wave of additional challenges. Weak defenders, therefore, have the incentive to bluff in early periods—fighting early entrants—and then acquiescing with increasingly probability as the number of remaining entrants declines. The third prediction is that entrants should base their strategy on information they can glean about the type of defender they are facing and the incentives this defender has to build a tough reputation over time. If a defender backed down in an earlier period, entrants know they are facing a weak or uncommitted opponent, and they should always enter. If the defender never

[19] This is especially true since defenders know that they will interact with each entrant only once.

[20] These predictions are based on a sequential equilibrium solution.

backed down, entrants should never enter in the early periods, since both weak and strong defenders will fight in early periods. They should then be more likely to enter during the middle and latter periods (knowing that weak defenders will be increasingly likely to back down at these times).

Importantly, allowing cheap talk alters none of these predictions. Sequential equilibria from the formal model indicate that the defender and the entrants should not change their behavior if cheap talk is possible.[21] This is because cheap talk does not alter the defender's or the entrant's payoffs in any way. Furthermore cheap talk provides no new information about the defender's type. Entrants should know that weak defenders will have an incentive to try and appear as if they were strong defenders, and hence issue threats.[22] Leaders can threaten to take action, but unless lying is costly to them, which it is not in our setup, it should not affect whether they fight or whether entrants choose to challenge.

The model, therefore, makes two predictions about the effect of cheap talk:

> H1: Defenders who issue a threat to fight will not deter any more entrants than those who do not issue a threat.

[21] To our knowledge no one has worked out the cheap talk version of the repeated entry deterrence game. We present our equilibrium analyses in an appendix designed for on-line posting. We show that in a single shot version of the game only a pooling equilibrium exists, and no separating or semi-separating equilibria exist. Our analysis shows that this holds in the repeated game setting as well.

[22] Although saying that you will not fight (which is off the equilibrium path) will almost certainly signal weakness.

H2: Defenders who threaten to fight should be no more likely to fight than those who did not.

Experimental Design

Again, our experimental design had two separate parts. The first did not allow communication, while the second did. In all cases, subjects were randomly assigned to two separate positions, entrants and defenders, which were referred to simply as first movers and second movers.[23] These neutral terms were used in order to avoid leading the subjects in any way. Defenders were also assigned a type, either weak or strong, which were called 'type 1' or 'type 2'. As indicated by the payoffs in Figure 1, strong

[23] Subjects were recruited through a university social science laboratory using an e-mail solicitation to all students who had signed up with the lab. Those who responded were accepted until all positions were filled. Subjects were only allowed to participate in the experiment one time. Students entered the laboratory one by one and were seated at computer workstations that were separated by pull out dividers to prevent interaction between subjects. Instructions were then read to all participants. During this process subjects were given the opportunity to make practice decisions and review a set of questions and answers about the experiment. Any questions from subjects were repeated and answered so that all subjects could hear. This ensured that all aspects of the experiment design were common-knowledge. Subjects were paid one by one at the end of the experiment with money earned in the experiment and a guaranteed $10 'show-up' fee. The experiment was programmed and conducted with the software z-Tree (Fischbacher 1999). Our design, instructions, and computer interface went through a lengthy piloting period in order to obtain the best possible experimental protocol, are provided for review and will be available on the author's webpage.

types prefer fighting after entry whereas weak types do not. Entrants were not told who was a strong or weak type - only that there was a one-third chance that any defender was strong. Each defender faced a sequence of eight entrants in a single repetition of the game, and this number was known to everyone. When an entrant was paired with a defender they played the game illustrated in Figure 1 a single time.[24] Entrants were also given information on how the defender played against all other previous entrants. If a previous entrant had chosen to challenge the defender, all subsequent entrants would see whether the defender had backed down or stayed tough. If an entrant chose not to challenge, no information about the defender's choice would be recorded.

The experiment proceeded as follows. Entrants faced the defenders sequentially. Within a pairing, entrants were asked to choose between entering the game and thus challenging the defender, or not entering. We elicited defender choices using the strategy method: defenders were asked to select a strategy based on what an entrant might do: 'if the first mover enters I will choose B1 or B2 (not fight or fight).[25] Each entrant

---

[24] All matching was entirely anonymous with subjects seated at separate partitioned computer terminals.

[25] We did this to observe the decision of a defender even when their opponent did not choose to enter. While in principle the mechanism of strategy solicitation can influence choices, there is considerable debate on this (Brandts and Charness 2000; McLeish and Oxoby 2004; Bosig et al. 2003). We note that behavior in our no communication treatment is very similar to that observed by (Bolton and Ockenfels 2007) whom elicited strategies sequentially. It is important to note that our design is not equivalent to using the

made one decision with no available history (in the first period), one decision with a previous period's history against a different defender (in the second period), and so on.[26] At the end of each repetition (after each entrant had played each defender once), subjects saw a screen with their decision history, the decisions of the subject they were paired with in each period, and their own payoffs.[27] Subjects knew that these payoffs would be translated into US dollars at the end of the experiment. Subjects then repeated the experiment.[28] Each repetition was done five times in order to account for the effects of learning and to generate sufficient data for the analysis.[29]

After completing the five repetitions, subjects were told that we were making a slight change in the experiment. We explained that defenders would now be able to

---

normal form version of the entry-deterrence game. Furthermore, we compare two treatments that used the same protocol to identify the effect of cheap talk.

[26] This design allowed us to keep all subjects engaged throughout the experiment, as well as maximize the amount of data we could collect within an experimental session.

[27] Payoffs to other players were not revealed in order to isolate the effect of learning across instead of within repetitions of the 8 period experimental round.

[28] Across repetitions of the experiment all positions (first mover/second mover) stayed the same, entrants were randomly assigned when they would move against each defender, and defender types (strong/weak) were randomly re-assigned according to the commonly known distribution of types.

[29] The precise number of repetitions was unknown to subjects; they were simply told that the experiment "may or may not be repeated" in order to limit attempts to build reputations across repetitions.

communicate to entrants whether they would fight or not. Defenders could do this by sending the following message through the computer: "if you choose enter, I will [fight, not fight].[30] This message was seen only by the immediate entrant and not by later entrants. Everything else in the experiment was the same as our baseline design and subjects were not told during the no-communication treatment that they would at some point have the option to communicate.[31] Hence our cheap talk experiment was run on a set of subjects with experience in the strategic environment of the repeated entry-deterrence game, but no prior experience with cheap talk. We identify the role of cheap talk by comparing behavior across the two treatments. An online appendix provides additional details and full subject instructions.

Results and Interpretation

Our goal in running the experiment was to collect data on how subjects played when no cheap talk was allowed versus how they played when it was.[32] We did this to

---

[30] Our experiment used neutral descriptions, and thus subjects actually chose between "I will (not) choose B1 if you choose A1". We did not allow subjects to not send a message.

[31] All subjects kept either their entrant or defender roles.

[32] Our empirical strategy for all of our hypotheses is to break defenders out by those who had already backed down and those that had not. We also break out entrants into those that face a defender who had not yet backed down, and those that faced a defender who had. We do this because the equilibrium model we discuss above makes this important distinction, and we do not want to conflate reputational effects with the effect of cheap talk. Next, we calculate either the mean rate of a behavior (e.g., taking the average of cases where entry=1 and no entry=0) and calculating test statistics using standard

answer two questions. First, would entrants be deterred by cheap talk threats or would they be equally likely to challenge in the face of a threat (H1)? Second, would defenders who issued a cheap talk threat be more likely to fight than those who did not, or would it have no effect at all (H2)? The results, which we discuss below, are striking.

Hypothesis 1: Entrants should not be deterred by cheap talk threats

Contrary to the implications of the formal model, cheap talk had a significant effect in our experiment. Potential entrants were more likely to be deterred when we allowed cheap talk than when we did not allow cheap talk. Figure 2 shows the entry rates of entrants across the two different experimental manipulations at each period in the game conditional on the defender not having back down in a previous time period. Later we consider cases where the defender has backed down. The figure reveals a dramatic difference in entry rates in the early periods - particularly in the first period. When communication is not allowed, fully 83% percent of entrants enter in the first round. However, when defenders are able to issue a verbal threat and chose to issue this threat, the high rates of entry in the first period disappear. When defenders engage in cheap talk, only 38% percent of potential challengers enter in the first round. That is a striking 45% percent decrease in the probability of entry and is significant at $p<.01$. This difference occurred despite the fact that the very same entrants were making these decisions.

The difference persists into the second period. As can been seen in Figure 2, there is a perceptible but not quite significant advantage to issuing a threat in the second time period. The small size of the difference in period two is, in part, due to combining

---

difference in means tests. Tests using differences in proportions produce nearly identical results.

two different groups of observations: entrants that faced a defender who had previously faced entry, and entrants who faced a defender that was not challenged in the first period. Once we consider this difference, it is clear that cheap talk still has a large effect in the second time period. In period two, when entrants faced a defender who had previously faced entry, the entry rate without communication was 25% but was a much lower 11% when a threat was sent. The effect of cheap talk was even larger when entrants faced a defender who had not been challenged in period one. In this case, the entry rate was 100% when communication was not possible and 52% when a threat was sent. Both differences were significant at $p<.01$. Importantly, since these results are conditioning on no previous backing down by the defender, the effect of cheap talk is a pure one. These results are unexpected. When entrants had little to no information about the type of defender they were facing, they were significantly more likely to be influenced by the messages they received even though these messages were costless.

Not surprisingly, the more information entrants were able to gather across periods, the less influential cheap talk became. As can be seen in Figure 2, costless threats did not continue to deter after the second time period, and by later rounds they actually caused entrants to be slightly *more* likely to challenge. We believe the influence of cheap talk declines over time because entrants obtain more reliable information about defender behavior by observing what the defender had done in the past against other entrants. Rather than having to rely solely on verbal promises, entrants could observe how a defender behaved against other challengers in previous rounds, and tie their strategies to these more dependable data.

The results so far suggest that cheap talk can work when little observable information is available on which entrants can make decisions. But what if entrants already have information that strongly suggests that the defender is weak? About 46% of the time, weak defenders chose to back down whereas less than 1% of strong defenders chose the not fight option. Entrants who observe a defender backing down can be fairly certain that they are facing a weak defender since strong defenders so rarely acquiesce. Can threats in this case—where presumably the defender's reputation for resolve has been lost— still make a difference?

Figure 3 suggests that they can. In Figure 3 we compare the relative probability of entry in the cheap talk versus the non cheap talk experiment for each time period in those cases where the defender had already backed down. Even in this extreme case, cheap talk still mattered. Entrants were still less likely to challenge in every period if a threat had been issued even if the defender had already revealed herself to be uncommitted. Due to the small number of cases, the difference is not significant in every individual time period but when we pool across all periods where there are observations in the threat and no-threat categories, we observe a highly significant difference between the cheap talk and non cheap talk versions of the game. Without communication, 95% of entrants chose to enter when their opponent had backed down previously, whereas only 85% entered after receiving a cheap talk threat ($t=2.56$, $p<.05$). This suggests that even a costless threat by a non-credible player has some deterrent value.

<u>Hypothesis 2: Defenders that threaten should not be more likely to fight.</u>

Costless verbal threats clearly influence whether entrants chose to fight or not in early periods of the game. But did it affect how defenders played? According to the logic of our formal model, defenders should not be more likely to fight after issuing a threat since there is no punishment for not following through. Did defenders who were allowed to threaten change their behavior in any way?

Our experiment reveals that weak defenders, at least early on, were more likely to fight if they said they would fight.[33] Figure 4 illustrates the rate of fighting across the two versions of the experiment by time period, and reveals that this difference is most pronounced in period 1 and then disappears after the first period.[34] Defenders are more likely to follow through in the very first period, and then taper off after that. Interestingly, this follow-through brings the behavior of weak defenders closer to the equilibrium predictions of the formal model with no communication.[35]

We were also able to see if the same subject changed his or her behavior when allowed to issue a threat. We found that 90% of our defenders increased the percentage

[33] We focus on the behavior of weak defenders because strong defenders should always fight (and almost always do). We exclude the first repetition of each treatment because behavior of defenders changed significantly in the no communication design from the first to the second repetition, where fight rates increased in all of our sessions. Including this repetition made the differences more significant because it decreased the proportion of defenders that fought in the no communication treatment.

[34] This is in part because the remaining weak types in the no communication design were the set of people that subsequently resisted in almost all of the remaining periods of play. These subjects were a subset of the subject sample that regularly played a much tougher strategy than other weak type defenders.

[35] Perhaps, as we will note below, after issuing a threat, defenders feel it would be dishonest to not follow through on that threat.

of times they fought in the first period if they were allowed to issue a threat.[36] Moreover, the change was usually large. On average, the same defenders were 16 percent more likely to fight when they had issued a threat than when they did not have the opportunity to issue a threat – a rate that is significantly different than 0 (t=2.05, p=.06).[37] Whether measured in the aggregate or at the individual level, cheap talk has a real effect on the behavior of individuals who engage in it.

These results indicate that cheap talk affects the behavior of defenders who have not previously backed down. But what happens after defenders have already signaled their type by backing down? Presumably, there is even less reason to follow through on threats in these cases.

Remarkably, costless communication still matters. As Figure 5 shows, cheap talk affects the behavior of defenders even after they had backed down. In each time period, the probability of fighting is higher when cheap talk is allowed than it is when cheap talk

---

[36] Here we calculated the total number of times that a subject chose to fight in the first period when they were assigned a weak defender role. We then divided this by the total number of times a subject was a weak defender in the first period (recall that *type* was randomly assigned and hence subjects might have different number of times that they played a weak defender role). This gives us a value between 0 and 1. We calculated this value separately for each treatment and each subject, and took the difference between these values for our subjects in the defender role.

[37] This suggests that across treatment, differences at the aggregate level move in the same direction as differences we observe within individual subjects; our aggregate differences are not driven by a single subject radically changing their behavior.

is not allowed. If we pool across periods 2-8, 13% of subjects in the cheap talk treatment decide to fight, whereas less than 5% percent decide to fight in the no communication treatment (t=-2.25, p<.05).[38],[39]

V. Explaining the Power of Cheap Talk

Our experiment investigated the role of cheap talk in a repeated entry-deterrence game and revealed that verbal communication can influence behavior in ways not captured by much of the formal and empirical literature in international relations. Verbal threats not only decreased an entrant's eagerness to challenge but also increased a defender's willingness to fight. Cheap talk may be costless, but it successfully deterred entrants and made defenders more willing to fight, at least in early rounds of our repeated game.

These findings bring us back to our original puzzle. If promises and threats are not worth the paper they are written on, as Samuel Goldwyn once said, why would anyone believe them, and even more puzzling, why would anyone follow through? One explanation relates to the willingness to lie. It is possible that some players do not engage in cheap talk because they prefer to be honest even if this means fewer payoffs as a result. If this were true, a separating equilibrium would emerge where "honest"

---

[38] There are no observations in period 1 because there were no previous periods in which a player could back down.

[39] Unpooling across periods radically reduces the sample size for which to conduct statistical tests. Thus it is not surprising that unpooling our analysis generates less significant test statistics for each period (results available from authors).

defenders would never threaten, and those who threatened would be more likely to follow through.

The idea that some people may be less willing to lie, and therefore, less willing to engage in cheap talk has found support in psychology studies. There is some evidence, for example, that men are more willing to lie than women (Keltikangas-Jarvinen and Lindeman 1997). There is also evidence that more extraverted individuals are more willing to deceive than more introverted ones (Weiss and Feldman 2006).[40] If it is true that certain "honest" individuals never engage in cheap talk, then the threats that are made are likely to be more credible and more effective as a result.[41]

The fact that costless threats affect entrant behavior even if defenders had already backed down lends more credence to this view. Under these circumstances, the defender has already revealed its type – weak or strong – and a costless signal at this point provides absolutely no information. But if that costless signal is sometimes an honest indication of what the defender is going to do, then even at this late stage of the game it can still provide information. Honest defenders who have backed down once may be

[40] Similarly, Majeski and Fricks (1995) found that some subjects who were more selfish were willing to use communication to exploit others, although the authors were unable to isolate how subjects did this.

[41] As we mentioned earlier, not all defenders chose to issue a threat. Eleven percent of weak defenders sent a signal that they would not fight. Fully 98% of the weak defenders who sent a signal that they would *not* fight ended up not choosing to fight. This is markedly lower rate of fighting than weak defenders who signaled that they would fight.

revealing whether they are likely to back down again. Thus, the signal may still be somewhat credible.

But what about the behavior of defenders? At first glance, it is not clear why weak defenders would be more likely to fight after issuing a threat. Since all communication is private (only the target receives the message) there is no reputational gain or loss for executing a threat. The defender also does not increase his or her payoffs by following through since the payoff structure is the same whether threats are issued or not.

We believe that honesty may play a role here as well. It is possible that some defenders gain psychological value from following through with their threats. Subjects who signal that they will fight and then choose not to follow through on that threat will have lied to their opponents. That may be easy for many of us to do, but it may be harder for others. To avoid the cognitive dissonance of speaking one way and acting another, some defenders may choose to follow through with their threats. There is, in fact, a large literature suggesting that cognitive dissonance is uncomfortable and that we often engage in complex actions to minimize this tension (Zimbardo and Leippe 1991; Festinger 1957). It is possible, therefore, that some defenders either feel obliged to follow through with a promise or prefer to follow through with that promise to avoid uncomfortable emotional feelings.

There is, however, a second possible explanation for at least some of the behavior of our subjects. One of the assumptions formal models make is that defenders and entrants interact under conditions of common knowledge. That means that everyone operates under the assumption that everyone understands the game and will play optimally as a result. It is possible, however, that common knowledge does not exist either in the laboratory or in the real world. Instead, a number of subjects may understand that some individuals will not "get" the game and will play poorly because of this. Mistakes may be made because some players misconstrue how the game should be played, misinterpret the instructions, or simply play irrationally for different idiosyncratic reasons. If this were true, cheap talk could be used by savvier subjects to signal to each other that they understand the game, ensuring that a higher proportion of efficient decisions are made.[42] In this case, it would once again be rational to respond to threats – however costless they may be.

This alternate explanation is especially plausible in the case of entrant behavior. Given that a threat is costless, a defender who threatens early in the game is playing exactly as one would expect him or her to play. Likewise, a player that does not issue a threat may be indicating that they do not fully understand the game. Thus, sending a threat or not sending a threat signals to the entrant something about the sophistication of their opponent. Knowing that your opponent is likely to play the game correctly by

[42] An argument similar to this was made by Vincent Crawford in an attempt to explain why deception might work (2003). Instead of having a distribution of honest types, as we argue may be the case, he considered that some people may be more easily 'fooled' than others. This creates a similar separation of types where rational players know that costless verbal communication will deceive at least those individuals who are less rational, making cheap talk sensible. This again suggests that there is a wider range of individuals than most models assume, and that certain types of individuals will behave quite differently from what existing models would expect.

fighting in these early rounds allows entrants to coordinate on the correct response, which is not to enter with a higher probability.

This common knowledge explanation, however, does not explain why defenders are especially apt to fight after issuing threats.[43] Once defenders have issued a threat, they have signaled how they plan to play and have gained all the value they can by creating common knowledge. If an entrant still chooses to enter, defenders garner no additional value by following through with their threats. In fact, under some circumstances, follow through can lead to diminished payoffs. Thus, there is no reason to expect the savviest players to be significantly more likely to follow through on their own threats. Future research will try to tease out what motivates individuals to follow up on their threats, perhaps by exploring models of cognitive consistency.

Conclusion

International relations has been skeptical about whether costless verbal communications have any influence on behavior despite the fact that state leaders engage in threats and promises all the time. In this paper, we put cheap talk to a particularly hard test: a situation where the preferences of the players are opposed and threats are private and costless. We expected that in an entry deterrence game where entrants were uncertain about whether defenders would fight, the use of costless threats would not change entrant and defender behavior in any way.

[43] It is, however, worth noting that this follow-through brings the behavior of weak defenders closer to the equilibrium predictions with no communication. The more consistent early period fighting of the defenders who issued a threat may be an indication that they understand the game better.

Controlling for confounding factors, our laboratory experiment revealed just the opposite. If a defender threatened to fight an entrant, that entrant was significantly less likely to challenge, and the defender was significantly more likely to fight especially in early rounds of the game. This occurred despite the fact that defenders suffered no punishment for failing to follow through with a threat; no additional entrants would know about the bluff and no other costs would be incurred. These findings bring academic research closer to what we have been observing in the real world. State leaders routinely issue verbal promises and threats, both publicly and privately, and sometimes these threats influence behavior.

Why did formal models miss these effects? We believe it has to do with at least one incorrect assumption underlying the models. Standard models assume that if there is a potential advantage to acting one way, all players will act to maximize that potential advantage. In other words, all players will act to maximize payoffs. Our laboratory experiment suggests that this assumption is not true, at least among the population of undergraduates we studied. Even among this relatively homogeneous subject pool, there appeared to be significantly more heterogeneity of preferences than the models predicted. The most important difference in terms of its effect on cheap talk was the existence of individuals who appeared to be averse to lying. The fact that a subset of subjects may have preferred honesty over money created an opening that may have allowed cheap talk to become influential. In the absence of these honest individuals no separating equilibrium would have emerged, and verbal threats would have provided no information. We also do not rule out other differences in terms of skill and the existence of individuals who did not fully comprehend the game. The prospect that some subjects could play the game poorly, may have allowed savvier defenders to signal their understanding of the game. Thus, the existence of honest and bumbling players in a given population may make even costless verbal messages rational and effective.

This does not mean that the same heterogeneity exists in the wider world, or amongst state leaders engaged in their own entry deterrence games. State leaders may be more willing to lie than the undergraduates we studied at Princeton. They may also be far more adept at navigating a complex strategic game. We strongly suspect, however, that the heterogeneity we found in the laboratory is not significantly different from the heterogeneity we are likely to find amongst leaders considering deterrence games in the wider world.

Thus, our research represents the beginning of a long agenda aimed at understanding why cheap talk matters and equally importantly, why cheap talk seems to be so influential under some circumstances and not others. The preceding paragraphs are an initial attempt to explain why our subjects behaved the way they did, but significantly more work needs to be done. We do not know, for example, what biases, beliefs or mental handicaps our subjects brought to the laboratory. We have some quotes from the subjects themselves, but their responses are unreliable and in need of additional analysis.

To this end there are a range of additional experimental designs that may help explain why cheap talk is more powerful than our existing theoretical approaches have predicted. Heterogeneity in individual behavior suggests that more could be done to understand differences across subjects. Additional information could be garnered by more extensive pre and post-experiment psychological batteries which might reveal correlations between behavior in the game and propensity to engage in other types of behavior (e.g., deceitful behavior). Akin to thinking about differences in individuals are differences across subject pools. We hope to extend our analyses to more targeted subject pools such as military officers and diplomatic officials (Mintz 2004). Finally, the ecological validity of the experiment could be increased (e.g., by making the decision context more concrete instead of abstractly described). All of these represent opportunities for future research that could be built on the results reported in this paper.

To date, laboratory experiments have rarely been used in international relations, especially with a game theoretic model to structure the design and empirical analysis. Our paper shows how this can be done in a substantively motivated way, with important results. Laboratory experiments can quickly and clearly reveal whether certain relationships hold, and if so under what conditions. This is a critical complement to much theoretical and empirical work in international relations. The experiment presented in this paper was a first step in explaining how and why individuals use verbal communication to influence each other's behavior in a repeated entry-deterrence game. We hope our findings encourage other researchers to theorize more deeply about cheap talk and to test various hypotheses about its effect on interpersonal and interstate relations.

Figure 2.

Challenger Behavior With and Without Cheap Talk: Entry Probability

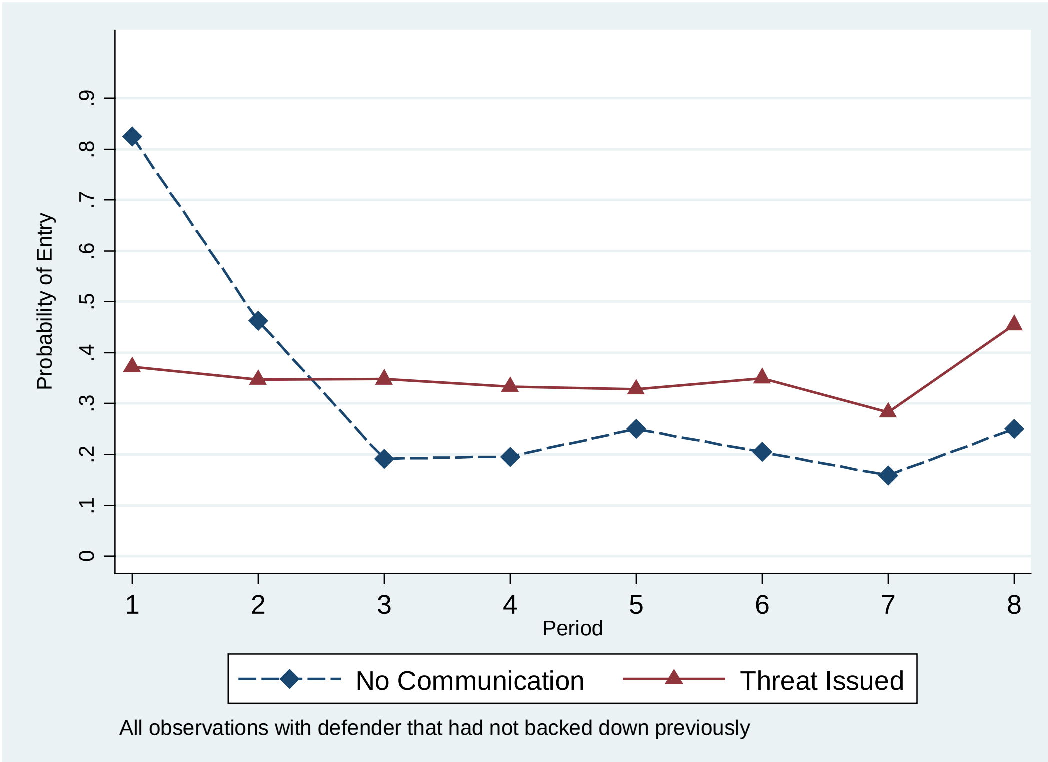


Figure 3

Challenger Behavior With and Without Cheap Talk:

Entry Probability after Defender Has Backed Down

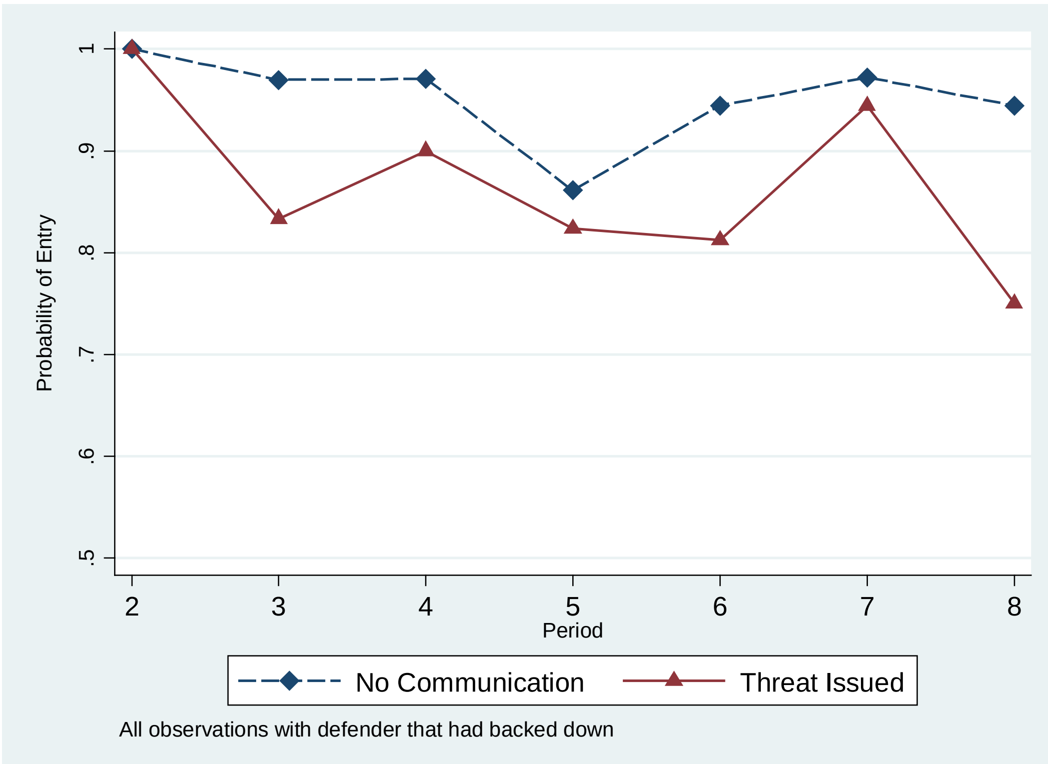


Figure 4

Defender Fighting Probability With and Without Cheap Talk and No Previous Backing Down

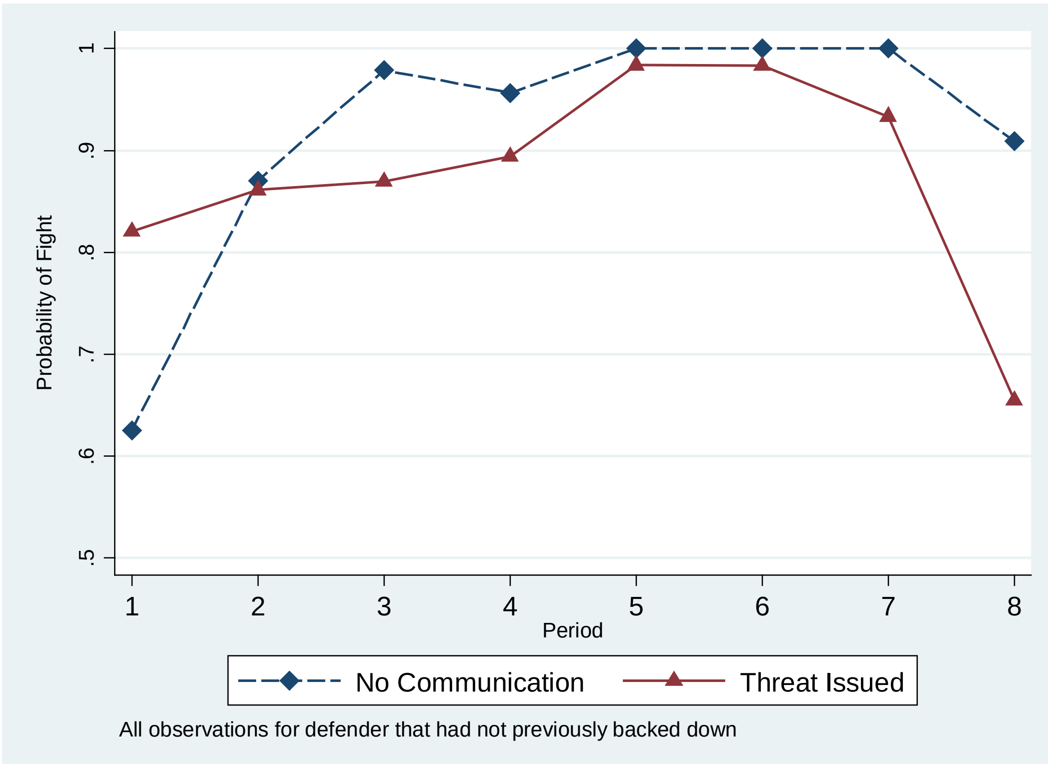
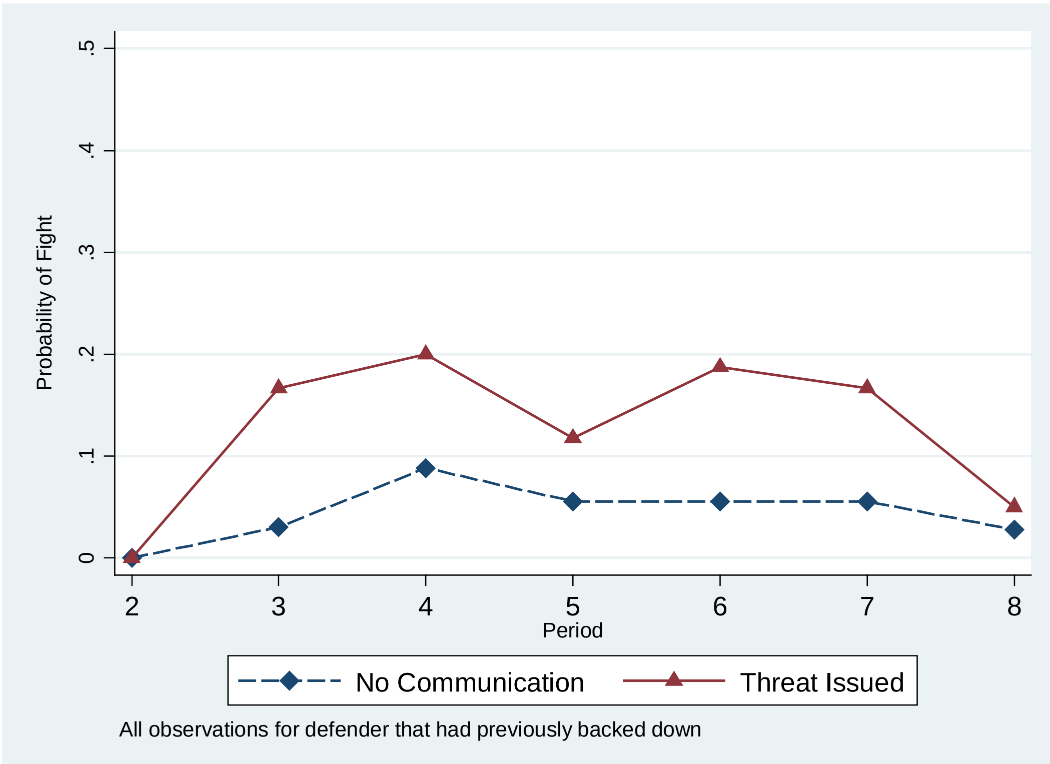


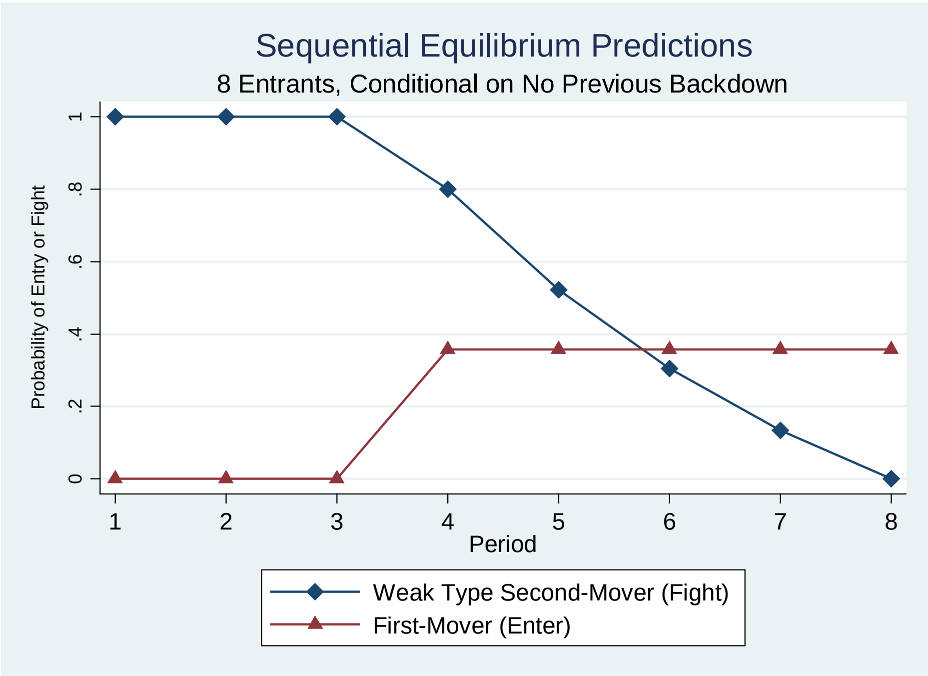
Figure 5

Defender Behavior With and Without Cheap Talk: Fighting Probability after Defender has Backed Down



APPENDIX

Given the number of entrants, the payoffs in the game, and the distribution of defender types (in our experiment 1/3rd were strong and 2/3rd were weak types) we can derive a sequential equilibrium as done in Jung et al. (Jung et al. 1994). Figure A graphs the probabilities of entry and fight where there has been no previous backing down. As we noted earlier, when a weak defender has previously backed down, the formal model indicates that challengers should always enter and the defender should always back down. Permitting cheap talk does not alter the predictions because only there is a strict incentive for all defenders to signal that they will fight and hence the signals are uninformative.



Bibliography

Bolton, Gary E., and Axel Ockenfels. 2007. "Information Externalities, Matching and Reputation Building - A Comment on Theory and an Experiment." http://ockenfels.uni-koeln.de/download/papers/bolton_ockenfels_information_externalities.pdf: University of Cologne Working Papers Series in Economics: 17.

Bosig, J., J. Weimann, and C.L. Yang. 2003. "The Hot versus Cold Effect in a Simple Bargaining Experiment." *Experimental Economics* 6:75-90.

Brandts, Jordi, and Gary Charness. 2000. "Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games." *Experimental Economics* 2 (3):227-38.

Cooper, R., D. DeJong, and R. Forsythe. 1989. "Communication in the Battle of the Sexes game: Some experimental results." *Rand Journal of Economics* 20:568-87.

Crawford, Joel. 1998. "A Survey of Experiments on Communication via Cheap Talk." *JOURNAL OF ECONOMIC THEORY* 78 (2):286-98.

Crawford, V., and J. Sobel. 1982. "Strategic information transmission." *Econometrica* 50:1431-52.

Crawford, Vincent P. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *The American Economic Review* 93 (1):133-49.

Croson, Rachel, Terry Boles, and J. Keith Murnighan. 2003. "Cheap talk in bargaining experiments: lying and threats in ultimatum games." *Journal of Economic Behavior and Organization* 51:143-59.

Davis, James W. 2000. *Threats and Promises: The Pursuit of International Influence*. Baltimore: Johns Hopkins University Press.

Farrell, Joseph. 1995. "Talk is Cheap." *The American Economic Review* 85 (2):186-90.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap Talk Can Matter in Bargaining." *JOURNAL OF ECONOMIC THEORY* 48:221-37.

Fearon, James. 1994. "Domestic political audiences and the escalation of international disputes." *American Political Science Review* 88:577-92.

———. 1995. "Rationalist Explanations for War." *International Organization* 49:379-414.

Festinger, L. 1957. *A theory of cognitive dissonance*. Evanston, IL: Peterson.

Finnemore, Martha, and Kathryn Sikkink. 2001. "Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics." *Annual Review of Political Science* 4:391-416.

Fischbacher, Urs. 1999. "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual, Working Paper Nr. 21." *Institute for Empirical Research in Economics, University of Zurich*.

Forsythe, R., J. Kennan, and B. Sopher. 1991. "An experimental analysis of strikes in bargaining games with one-sided private information." *American Economic Review* 81:253-78.

Foster, Dennis M. 2006. "Invitation to Struggle? The use of force against “legislatively vulnerable” American Presidents." *Journal of International Studies Quarterly* 50 (2):421-44.

Guisinger, Alexandra, and Alastair Smith. 2002. "Honest threats: The interaction of reputation and political institutions in international crises." *Journal of Conflict Resolution* 46:175-200.
Isaac, R., and J. Walker. 1988. "Communication and free-riding behavior: The voluntary contribution mechanism." *Economic Inquiry* 26:585-605.
Jung, Yun Joo, John H. Kagel, and Dan Levin. 1994. "On the existence of predatory pricing: an experimental study of reputation and entry deterrence in the chain-store game." *The Rand Journal of Economics* 25 (1):72-93.
Keltikangas-Jarvinen, Liisa, and Marjaana Lindeman. 1997. "Evaluation of Theft, Lying, and Fighting in Adolescence." *Journal of Youth and Adolescence* 26 (4).
Kreps, David, and Robert Wilson. 1982. "Reputation and Imperfect Information." *JOURNAL OF ECONOMIC THEORY* 27:253-79.
Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47 (4):597-611.
Majeski, Stephen, and Shane Fricks. 1995. "Conflict and Cooperation in International Relations." *Journal of Conflict Resolution* 39:622.
McDermott, Rose, Jonathan Cowden, and Cheryl Koopman. 2002. " Framing, Uncertainty, and Hostile Communications in a Crisis Experiment." *Political Psychology* 23 (1):133-49.
McLeish, K., and R. Oxoby. 2004. "Specific Decision and Strategy Vector Methods in Ulimatum Bargaining: Evidence on the Strength of Other-Regarding Behavior." *Economics Letters* 84:399-405.
Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *JOURNAL OF ECONOMIC THEORY* 27:280-312.
Mintz, Alex. 2004. " Foreign Policy Decision Making in Familiar and Unfamiliar Settings: An Experimental Study of High-Ranking Military Officers." *Journal of Conflict Resolution* 48 (91):91-104.
Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants With and Without A Sword: Self-Governance is Possible." *American Political Science Review* 86 (2):404-17.
Palfrey, T., and H. Rosenthal. 1991. "Testing for effects of cheap talk in a public goods game with private information." *Games and Economic Behavior* 3:183-220.
Ramsay, Kristopher. 2004. "Politics at the Water’s Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48:459-86.
Risse, Thomas. 2000. ""Let's Argue!": Communicative Action in World Politics." *International Organization* 54 (1):1-39.
Sartori, Anne. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56:121-50.
———. 2005. *Deterrence By Diplomacy*. Princeton, NJ: Princeton University Press.
Schrodt, Philip A. 1993. "Foreign Policy Analysis: Continuity and Change in Its Second Generation." In *In Event Data in Foreign Policy Analysis*, ed. P. J. Haney, L. Neack and J. A. K. Hey: Prentice Hall.
Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92 (4):829-44.
Sobel, Joel. 1985. "A Theory of Credibility." *The Review of Economic Studies* 52 (4):557-73.

Sundali, James A, and Darryl A Seale. 2004. "The Value of Cheap-talk and Costly Signals in Coordinating Market Entry Decisions." *Journal of Business Strategies* 21 (1):69-94.

Thyne, Clayton L. 2006. "Cheap Signals with Costly Consequences: The Effect of Interstate Relations on Civil War." *Journal of Conflict Resolution* 50:937-61.

Trager, Robert F. 2009. "Multi-Dimensional Diplomacy." *working paper*.

Walter, Barbara F. 2006. "Building Reputation: Why Governments Fight Some Separatists but Not Others." *American Journal of Political Science* 50 (2):313-30.

Weiss, Brent, and Robert Feldman. 2006. "Looking Good and Lying to Do It: Deception as an Impression Management Strategy in Job Interviews." *Journal of Applied Social Psychology* 36 (4):1070-86.

Wilson, Rick K., and Jane Sell. 1997. ""Liar, Liar... ": Cheap Talk and Reputation in Repeated Public Goods Settings." *Journal of Conflict Resolution* 41:695-717.

Zimbardo, P. G., and M. Leippe. 1991. *The psychology of attitude change and social influence*. 3rd ed. New York: McGraw-Hill.

Instructional Materials for Subjects

During the instruction period we read the following script and provided subjects with two worksheets. A Powerpoint presentation was also used to supplement the script (provided at end).

Experiment Instructions

Thank you for agreeing to participate in this research experiment on group decision making. During the experiment we require your complete, undistracted attention. You may not chat with other students, or engage in other distracting activities, such as using your phone or headphones, reading books, etc. Please turn all cell phones to silent.

For your participation, you will be paid in cash, at the end of the experiment. Different participants may earn different amounts. You will be paid privately, and are under no obligation to tell others how much you earned. What you earn depends on your decisions and the decisions of others. It is very important that you follow the instructions closely. All participants in this experiment receive the exact same set of the following instructions.

The entire experiment will take place through computer terminals, and all interaction between each of you will take place through the computers. It is important that you not talk or in any way try to communicate with other subjects during the experiments. If you disobey the rules, you will be asked to leave the experiment.

We will start with a brief instruction period. If you have any questions during this period, raise your hand and your question will be answered so that everyone can hear. If any difficulties arise after the experiment has begun, raise your hand, and an experimenter will come and assist you.

Instructions

Subjects will be split into two groups, a ‘first mover group and a ‘second mover’ group. Your assignment will be the same for the whole session. Whether you are a first mover or a second mover is determined randomly and shown on your computer screen once the experiment starts.

<u>The decision situation</u>

The experiment is divided into eight periods. In every experiment round each of the second-movers is paired with eight different first-movers. An experiment round ends when the second mover has been paired with each of these first movers once. No one will play the same person twice in an experiment round.

The experiment begins with first movers choosing between one of two alternatives. The decision situation is projected at the front of the room. These alternatives are labeled A1 and A2 respectively. Choosing 'A1' produces an amount of points that depends on how second mover subjects respond. Note that if A2 is chosen, the second mover's choice does not affect the outcome. The second mover they are paired with then chooses 'if A1 was chosen, I will choose B1' or 'if A1 was chosen, I will choose B2' on a similar screen.

Comment [d1]: Screen 2

This process repeats until everyone has been paired with everyone *once*. Thus, each second mover will encounter a sequence of 8 different first movers in a round. Each first mover will play each second mover, but each time at a later period in the round.

To illustrate how you will be paired with other subjects and to show you how to read information provided on your screen, we will take you through an example round of the experiment. Please follow our directions exactly. Please click on the icon titled zLeaf on your desktop. On the first screen you are told whether you are a first mover or second mover. If you are a second mover you are also told whether you are type 1 or type 2. Will explain the difference in a moment. Please click OK. On the next screen you are able to make a decision. If you are a first mover you are paired with a single second mover, and vice versa. The left hand side of the screen will report information about the second mover in the pairing. This information will be the choice of previous first movers they are paired with, and the choice of the second mover if A1 is chosen.

Comment [d2]: Screen 3

This screen does not yet have any information. An example of what a first mover will see in the first period is projected at the front of the room. Now, we are projecting what a second mover will see in the first period.

Comment [d3]: Screen 4

Comment [d4]: Screen 5

If you are in the first row of seats, choose A1 if you are a first mover and B1 if you are a second mover. If you are in the second row, choose A2 if you are a first mover and B2 if you are a second mover. Please remember to hit ok after your decision. If you are the third row of seats choose A1 if you are a first mover and B1 if you are a second mover. The next screen you see is slightly different that the previous screen. If you are a first mover, you are now paired with a different second mover. The information you see on the left side of the screen is information about what this second mover faced in the first period. Unlike the decision made in the first period, all subjects have a record of what the second mover faced in the previous period. If the first mover they were paired with chose A1, you will see how the second mover responded. If the first mover they were paired with chose A2, you will not see the decision of the second mover. If you are a second mover, you see exactly the same information that the person you are currently paired with sees. This information is your own history: what the first mover you faced in the first period did, and what your response was if they chose A1.

The screen at the front of the room shows what a first mover might see in the second period. In this case, they are paired with a second mover whose choice in the first period did not matter, because the first mover they were paired with chose A2. The next screen

Comment [d5]: Screen 6

Comment [d6]: Screen 7

shows what a second mover might face in the second round. This shows that the first mover they were paired with chose A2, and thus their choice is not recorded. Are there any questions?

If you are in the first row, choose A2 if you are a first mover and B2 if you are a second mover. If you are in the second row, choose A1 if you are a first mover and B1 if you are a second mover. If you are in the third row, choose A1 if you are a first mover and B1 if you are a second mover. Please hit 'ok'. If you are a first mover, you are now paired with a different second mover. The information you see on the left side of the screen is information about what this second mover faced in the first period and second period. If you are a second mover, you see exactly the same information that the person you are currently paired with sees. This information is your own history: what the first mover you faced in the first period did, what a different first mover did in the second period, and what your responses were if they chose A1. The screen at the front of the room shows what a first mover might see in the third period. In this case, they are paired with a second mover whose first period choice was B2 in response to a first mover that chose A1. In the second period, this same second mover faced a first mover that chose A2, and so the second mover choice was not recorded.

Comment [d7]: Screen 8

If you are in the first row, choose A1 if you are a first mover and B2 if you are a second mover. If you are in the second row, choose A2 if you are a first mover and B1 if you are a second mover. If you are in the third row, choose A2 if you are a first mover and B2 if you are a second mover. Please hit 'ok'. You now see the screen for the fourth period. The screen projected at the front of the room shows the information a second mover might see. In the first period, their first mover chose A2, and thus their response is not recorded. In the second period their first mover (a different person) chose A1 and they responded B1. In the third period their first mover chose A2 and thus the second mover choice is not recorded. Please take a moment to review the information on the left side of your screen. Remember, all of the information reflects what we told everyone to do. In the experiment, everyone will be able to make his or her own choices.

Comment [d8]: Screen 9

If you are in the first row, choose A2 if you are a first mover and B1 if you are a second mover. If you are in the second row, choose A1 if you are a first mover and B2 if you are a second mover. If you are in the third row, choose A1 if you are a first mover and B1 if you are a second mover. Please hit 'ok'.

This process would continue for eight periods, until all first movers and second movers have been paired once. At the end of the eighth period you will see the points you earned in the round.

Instructions for points

You have been provided a diagram that lists how points are earned by subjects. Please turn to the side that says 'first mover subjects'.

If you are assigned a first mover role, you should use this side. Please focus your attention on screen at the front of the room. First movers can choose A1 or A2. If the first mover chooses A2, the FM gets 95 and the second mover gets 300 pts. If the first mover chooses A1, then the points it earns will depend on what the second mover does. If the second mover chooses B1, the first mover earns 150 pts. If the second mover chooses B2, the first mover earns 80 pts.

Comment [d9]: Screen 10

Now let's talk about the second mover. The second mover gets the most points if the first mover chooses A2 – 300 points. But if the first mover chooses A1, then the second mover's points depend on whether the second mover is a type 1 or type 2. Approximately 2/3rds of second mover's will be type 1, and approximately 1/3 will be type 2. Type 1 second movers get more points for choosing B1 (160 pts) than choosing B2 (70 pts). This scenario is listed at the top of the page and is projected on the screen at the front of the room. Type 2 second movers get more points from choosing B2 (160) than B1 (70). This scenario is listed at the bottom of the page and is now projected on the screen at the front of the room.

Comment [d10]: Screen 11

Comment [d11]: Screen 12

Are there any questions?

Please turn to the side that says 'second mover subjects'. If you are a second mover, you should use this side. The situation is the same as just we just described, but is presented from the perspective of the second mover. If you are a type 1 second mover and A1 has been chosen, you receive more points from choosing B1 than B2. If you are a type 2 second mover and A1 has been chosen, you receive more points from choosing B2 than B1. All second movers earn the most amount of points by the first mover choosing A2. All first movers earn the most amount of points by choosing A1 and the second mover choosing B1. The next highest amount of points comes from choosing A2. First movers receive the least amount of points from choosing A1, and then the second mover choose B2.

Comment [d12]: Careen 13

Are there any questions?

The experiment may, or may not, be repeated several times. If the experiment is repeated, you will keep the same first mover/second mover assignment, and second movers will keep their same type.

In this experiment, you will be playing for real money. All earnings are expressed in terms of points. Points will be converted to dollars at the rate of $2.00 per 1000 points. If the experiment is repeated, this conversion rate will be adjusted. For example, if the experiment is repeated twice, you will earn $2.00 per 2000 points. In addition, all subjects are paid a $10 show-up fee. Your earnings do not depend on the earnings of other subjects in your sub-group. For example, if you are a first mover, your earnings do not depend on the earnings of other first movers.

At the end of the experiment you will fill out a brief feedback form. When you have completed the feedback form please wait. We will notify you when we are ready for you to exit the room.

We have provided a sheet of paper with the basic information you need to know. On the opposite side of the paper we have listed several questions with answers. Please take a moment to read through this material.

Are there any questions before we begin the experiment?

We will now begin. If you are assigned a second mover role, please record whether you are of type 1 or type 2 at the top of your worksheet.

After completing 5 repetitions of the experiment, subjects were then told:

We will now change the experiment slightly. Each time a first mover is paired with a second mover the second mover will have an opportunity to issue a statement to the first mover. This statement will be "if you choose A1 then B1 will be chosen" or "if you choose A2 then B2 will be chosen". The first mover will see this statement at the top right hand side of the screen. The first mover will also see the second mover's history in previous periods with other first movers, as was the case in the first experiment. As in the first experiment, the second mover will be asked to make an actual decision between choosing B1 if A1 was chosen, or B2 if A2 is chosen. All other features of the experiment are exactly the same as in the first experiment.

To illustrate what this looks like, please direct your attention to the screen at the front of the room.

Comment [d13]: Screen 1

This screen allows a second mover to send a message to the first mover they are paired with, saying that if A1 is chosen they will choose B1 or B2.

The next screen is what a first mover might see in the first period. Here, the second mover they are paired with has sent a message saying that if A1 is chosen, then B2 will be chosen. This message is in the top right hand side of the screen. As in the first experiment, the history screen on the left hand side is blank because this is the first period.

Comment [d14]: Screen 2

The screen you now see is what a first mover might see in the third period. Here, the first mover sees what the second mover they are paired with faced in the previous two periods. The first mover also sees the message that the second mover has sent them. In this case, the first mover is being told that if A1 is chosen, then B2 will be chosen.

Comment [d15]: Screen 3

To review, second movers will now have an opportunity to send a message to the first mover they are paired with stating what they will do if A1 is chosen. Second movers must also choose what they will do if A1 is chosen. The statement and action do not need to be identical. First movers will see the statement made by the first mover and will select between A1 and A2 as before. First movers do not see a history of statements by the second mover they are paired with, only the second mover's actions and the actions of the first mover they were paired with. All other features of the experiment remain the same. 1/3rd of second movers will be type II second movers who prefer choosing B2 to B1 if A1 is chosen.

HANDOUTS GIVEN TO SUBJECTS

Basic Features of the Experiment

- You will be divided into two groups: (1) First Movers and (2) Second Movers
- Second Movers come in two types:
    - 2/3 are type 1 (get more points for choosing B1 than B2)
    - 1/3 are type 2 (get more points for choosing B2 than B1)
- The Experiment will consist of eight periods. We may or may not repeat the experiment.
- A single period = one Second Mover paired with one First Mover
- An experiment round = one second mover paired with eight different first movers
- Information will be available on the history of decisions within a round: First Movers see the choices previous first movers made, and what the Second Mover did if A1 was chosen, for the second mover they are currently paired with. Second movers see their 'own' history, which is their own decisions and the decisions made by each first mover they were paired with.
- Points: will be determined by the choice you make and the choice your partner makes. Points are converted into cash

Review questions

1) What does it mean when you see a blank information screen?

Answer: You are in the first period and do not have any information about how second movers decided in previous periods.

2) If you are a first mover, does the amount of points you earn depend on the amount of points earned by other first movers?

Answer: No. The points you earn as a first mover only depends on your own choices, and the choice made by second movers you are paired with in a given period.

3) In the experiment there will be 8 first movers and 8 second movers. Of the 8 second movers, *approximately* how many of them will be a type 2 second mover.

Answer: 1/3, or approximately 2 or 3. This assignment of 'type' is determined by the computer randomly. On average in this experiment, 4 out of the 12 second movers will be a type 2 second mover, who gets more points for choosing B2 than B1.

4) If you are a first mover, what does the information on the left side of the screen mean?

Answer: this information is what the second mover you are currently paired with faced in earlier periods in the current round. It lists the first mover's decision, and the second mover's response if A1 was chosen.

5) If you are a second mover, will the first mover you are paired with in the third period be able to see what you did in the first period?

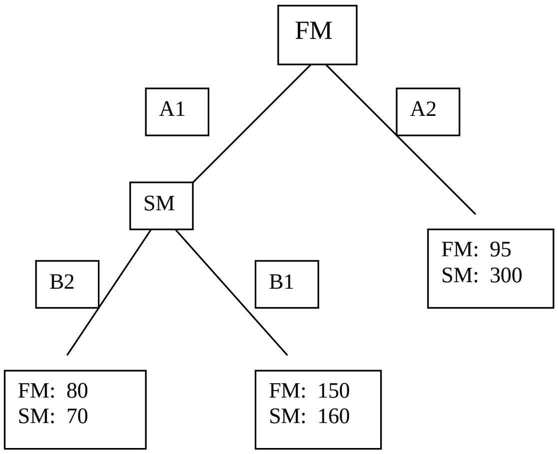
Answer: Yes. As long as the first mover chose A1 in the first period, all following first movers (7 in a round) you are paired with will see your choice from the first period.

USE THIS SIDE IF YOU ARE A SECOND MOVER

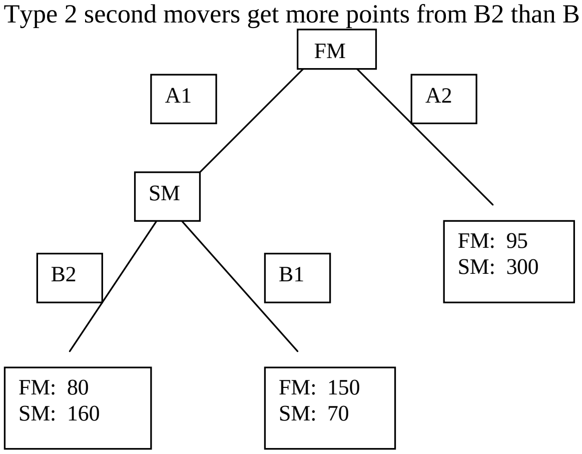
I am a type ___ second mover

<u>If you are a type 1 second mover</u>:

If the first mover chooses A2 you get 300 pts and they get 95 pts
If the first mover chooses A1 and you choose B1 you get 160 pts and they get 150 pts
If the first mover chooses A1 and you choose B2 you get 70 pt and they get 80 pts
Type 1 second movers get more points from B1 than B2 *if* A1 is chosen



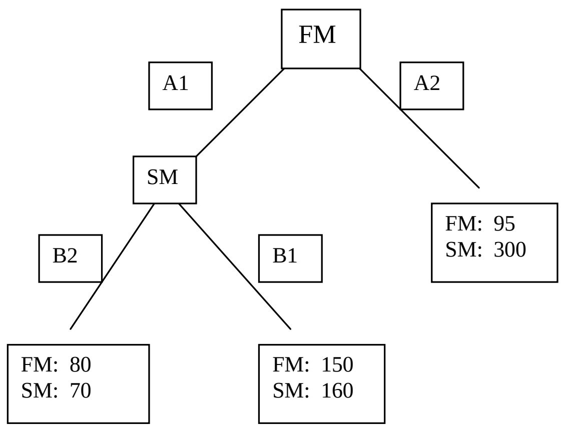
<u>If you are a type 2 second mover</u>:
If the first mover chooses A2 you get 300 pts and they get 95 pts
If the first mover chooses A1 and you choose B1 you get 70 pts and they get 150 pts
If the first mover chooses A1 and you choose B2 you get 160 pt and they get 80 pts
Type 2 second movers get more points from B2 than B1 *if* A1 is chosen



USE THIS SIDE IF YOU ARE A FIRST MOVER

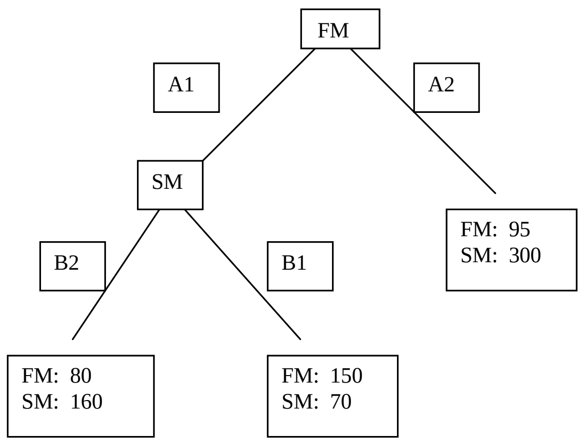
<u>If you face a type 1 second mover (probability = 2/3)</u>:

If you choose A2 you get 95 pts and they get 300 pts
If you choose A1 and they choose B1 you get 150 pts and they get 160 pts
If you choose A1 and they choose B2 you get 80 pts and they get 70 pts
Type 1 second movers get more points from B1 than B2 *if* A1 is chosen



<u>If you face a type 2 second mover (probability=1/3)</u>:

If you choose A2 you get 95 pts and they 300 pts
If you choose A1 and they choose B1 you get 150 pts and they get 70 pts
If you choose A1 and they choose B2 you get 80 pts and they get 160 pts
Type 2 second movers get more points from B2 than B1 *if* A1 is chosen



Supplementary Equilibrium Analyses for
Does Cheap Talk Matter? An Experimental Analysis

This memo sketches out the equilibrium properties for several models with a one-sided incomplete information entry deterrance stage game. For pedagogical purposes we consider four variations of the model along whether 1) the game was a single shot game or was repeated and 2) whether pre-play communication (cheap talk) is possible. Algebraic steps are included. This memo will be provided on the author's website alongside supplementary materials for the experimental sessions.

Figure 1 gives the standard entry deterrence game with entrants that have incomplete information about whether they face a tough opponent (committed) with probability P or a weak opponent (uncommitted) with probability $1 - P$.

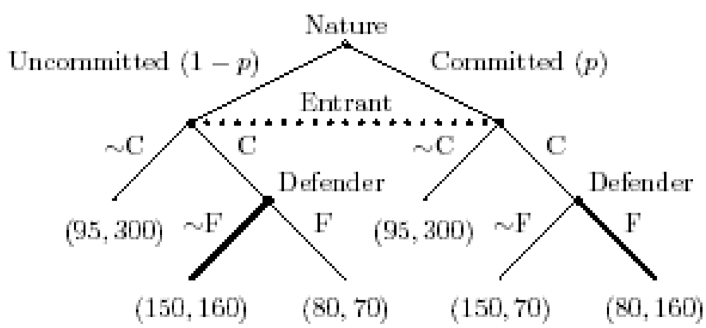


Figure 1: The Structure of a Single-Shot Play

1 No Signalling

1.1 Single shot play

Consider first a single shot version of the game in figure 1.

Proposition 1 *There is a unique equilibrium of the single shot game with prior entrant beliefs that the defender is strong of $P = \frac{1}{3}$ where the entrant enters with probability one, weak defenders back down, and strong defenders fight.*

Proof: In a single shot version of this game the equilibrium depends only on the prior, p, of facing a strong type. The expected utility of entry is $EU(Enter) = (1-p)150 + 80p = 150 - 70p$. The expected utility of not entering it is 95. The entrant will choose to enter when $150 - 70p > 95$, or when $p < \frac{55}{70}$. In the game we consider $p = \frac{1}{3}$ and hence entrants always enter and defender's fight if they are strong and don't fight if they are weak.

1.2 Repeat Play

Consider now a repeated game with the following features. A single defender faces a series of $n = 8$ entrants. Each entrant is uncertain about the defender's type, but later defenders can observe the choices of the defender in earlier periods if the defender faced entry in an earlier period. The sequential equilibrium solution, as described below, is given by Jung et al. 1994. I reiterate their logic using their notation.

Proposition 2 *A sequential equilibrium exists where weak defenders fight with probability 1 in the first three periods and then play a strictly declining mixed strategy in the remaining periods. Entrants do not enter in the first three periods and then enter with probability .357 in the remaining periods. Finally, off-path beliefs are restricted such that observing backing down leads to a posterior of* $P = 0$ *(Jung et al 1994).*

Index periods such that the first period is period n, the second period is $n-1$, third period $n-2$,...1. In the final period the last entrant will have some posterior belief, P_1, about whether they face a strong defender. They enter if $80P_1 + 150(1 - P_1) > 95$. I.e., they enter if $P_1 < \frac{55}{70} = Q$. In the final period a weak type will not fight entry because there are no reputational benefits. In period 2, the second to last period, the defender can play a mixed-strategy, where they fight with probability T_2. Hence the expected utility to the second to last entrant for entry is $80P_2 + 80(1-P_2)T_2 + 150(1-T_2)(1-P_2)$. They will choose entry when $80P_2 + 80(1-P_2)T_2 + 150(1-T_2)(1-P_2) > 95$.

Now consider how Bayes rule is used to update beliefs in the last period. We want

$$\begin{aligned} P_1 &= \Pr(Strong|Fight) \\ &= \frac{\Pr(Fight|Strong)\Pr(Strong)}{\Pr(Fight)} \\ &= \frac{\Pr(Fight|Strong)\Pr(Strong)}{\Pr(Fight|Strong)\Pr(Strong) + \Pr(Fight|\tilde{\ }Strong)\Pr(\tilde{\ }Strong)} \\ &= \frac{1P_2}{P_2 + T_2(1-P_2)}. \end{aligned}$$

Now we want to know when P_1 is lower than the threshold for entry in the final state. I.e., when does $Q < \frac{P_2}{P_2+T_2(1-P_2)}$? Solving for T_2:

$$
\begin{aligned}
Q &< \frac{P_2}{P_2+T_2(1-P_2)} \\
Q(P_2+T_2(1-P_2)) &< P_2 \\
QP_2+QT_2-QT_2P_2 &< P_2 \\
QT_2(1-P_2) &< P_2(1-Q) \\
T_2 &< \frac{P_2(1-Q)}{Q(1-P_2)}
\end{aligned}
$$

In a mixed strategy equilibrium the defender will choose T_2 such that this expression is an equality. Substituting this expression in to the above for whether or not the entrant enters in the second to last period, $80P_2 + 80(1-P_2)T_2 + 150(1-T_2)(1-P_2)$, we obtain the following.

$$
\begin{aligned}
80P_2+80(1-P_2)T_2+150(1-T_2)(1-P_2) &> 95 \\
80P_2+80(1-P_2)\frac{P_2(1-Q)}{Q(1-P_2)}+150(1-\frac{P_2(1-Q)}{Q(1-P_2)})(1-P_2) &> 95 \\
80P_2+80\frac{P_2(1-Q)}{Q}+150(1-P_2-\frac{P_2(1-Q)}{Q(1-P_2)}+\frac{P_2^2(1-Q)}{Q}) &> 95 \\
80(\frac{P_2Q+P_2-P_2Q}{Q})+150(\frac{Q-P_2}{Q}) &> 95 \\
\frac{80P_2}{Q}+150-\frac{150P_2}{Q} &> 95 \\
\frac{-70P_2}{Q} &> -55 \\
P_2 &< \frac{55}{70}Q \\
P_2 &< Q^2
\end{aligned}
$$

Similarly, the entry constraint at period n is $P_n < Q^n$.

Now consider the weak defender's mixed strategy in the nth period. From above we had $T_2 = \frac{P_2(1-Q)}{Q(1-P_2)} = \frac{P_2(1-Q^{2-1})}{Q^{2-1}(1-P_2)}$. Hence our expression for the nth period is simply $T_n = \frac{P_n(1-Q^{n-1})}{Q^{n-1}(1-P_n)}$ such that $\frac{P_n(1-Q^{n-1})}{Q^{n-1}(1-P_n)} < 1$ (because will be mixing). Hence $T_n = \min[1, \frac{P_n(1-Q^{n-1})}{Q^{n-1}(1-P_n)}]$ (because a mixture can't invovle a probability over 1).

In the first period there is the prior p that the defender is strong. If $P_n > Q^n$ then the entrant should not enter. In such a case, $\frac{P_n(1-Q^{n-1})}{Q^{n-1}(1-P_n)} > 1$ and $T_n = 1$. Thus the weak defender will fight when $P_n > Q^n$. During these early periods if a defender does not fight then off the equilibrium path beliefs are assumed to be that they are strong with probability 0. E.g., if no fighting happens in period 1, $P_{n-1} = 0$, and all subsequent entrants enter with probability 1.

Now consider when $P_n \leq Q^n$, this will generate the mixed-strategy phase of play. In the mixed strategy period entrants must be entering with a probability, O_n, that makes weak defenders indifferent. We begin with the next to last stage, as in the final stage defenders never fight. The defender must be made indifferent between playing some mixture T_2 and backing down for sure in both periods. Assuming entry in the second to last period (and hence the first term in the brackets is 70), their expected utility from the mixture is $T_2[70+160O_1+300(1-O_1)]+[1-T_2](160+160)$. Their expected utility of backing down in the second to last period is simply $160+160O_1+(1-O_1)160 = 320$. Hence we have the indifference condition:

$$\begin{aligned} T_2[70+160O_1+300(1-O_1)]+[1-T_2](320) &= 320 \\ 70T_2+160O_1T_2+300T_2-300O_1T_2+320-320T_2 &= 320 \\ 370T_2-320T_2+160O_1T_2-300O_1T_2+320 &= 320 \\ 320+T_2[(370-320)-O_1(300-160)] &= 320 \end{aligned}$$

Hence to establish indifference we must have that $T_2[(370-320)-O_1(300-160)] = 0$, which occurs when $O_1 = \frac{(370-320)}{(300-160)} = .357$.

It is important to note that the final period play by the entrant is NOT the same as single shot play (where $O_1 = 1$). This is because the entrant's *posterior* belief in the final period is updated to the point their posterior in the last period is exactly equal to the threshold, $P_1 = \frac{P_2}{P_2+T_2(1-P_2)} = Q$, and hence they are indifferent between entry and not entry and are playing a mixed strategy.

The entrant will choose their mixture in the second to last period in a similar manner. They weigh what they would get if they played a mixture in the third to last period to what they would get if they started backing down. With some algebra we see that $O_2 = \frac{(370-320)}{(300-160)} = .357$. By this logic, in equilibrium and during the mixed strategy phase, the entrant enters with a constant probability.

Finally, if there has been no backing down by the defender, play is in the mixed strategy phase, and the previous entrant did not enter, then the defender does not have the opportunity to build their reputation by resisting entry. Hence we have $P_n = Q_n < Q^{n-1}$. This constraint will always hold and thus there will be entry in period $n-1$. The empirical analysis in the paper ignores this part of the equilibrium, but additional analyses suggested that had little bearing on our results.

A key assumption in this sequential equilibrium is that if an entrant ever observes that their opponent has backed down, then their belief that the defender is strong is restricted to be 0. This of course should never happen in the pure strategy phase, and hence the reputational dyanamic implicit in the model turns on the existence of a "crazy" type.

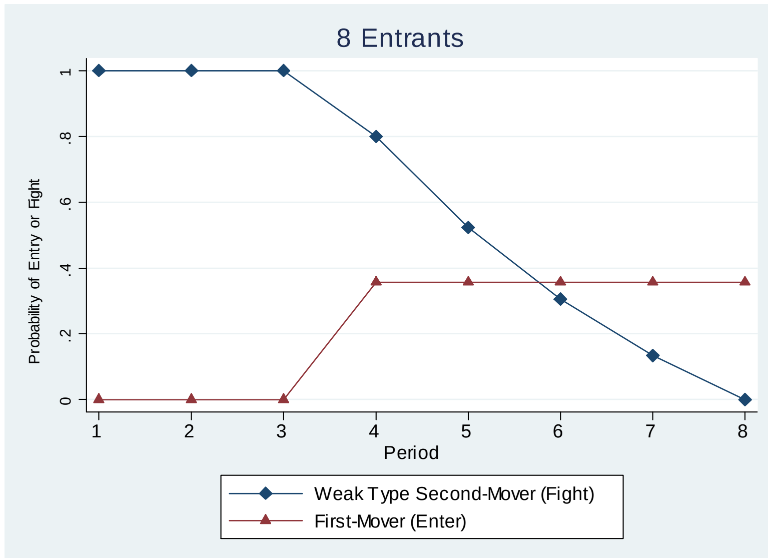


Figure 2: Sequential Equilibrium Predictions With Repeated Play

2 Signalling

In this section we modify the game to include an opportunity for the defender to issue a signal prior to strategy choices. All other features remain the same.

2.1 Single Shot Play

Prior to the entrant moving (but after "Nature" assigns type) the Defender chooses to send a signal Fight=θ_F or Not Fight= θ_{NF}. Hence the signal space is $\Theta = (\theta_F, \theta_{NF})$. Then Entrant chooses E or $\tilde{}E$ and Defender chooses F or $\tilde{}F$ if then Entrant chose E. The defender's type is unknown to the entrants, but they have a common prior $P = \frac{1}{3}$ that the defender is the tough type and $1 - P$ that they are the weak type. Now consider a single shot of the game but

with signalling. We investigate three classes of equilibria, separating, pooling, and semi-separating. Of course, babbling equilibria always exist (where pooling is one type of babbling equilibrium (see discussion in pooling section).

Separating

Proposition 3 *No separating equilibrium exists.*

Proof: Assume a separating equilibrium exists. In a separating equilibrium weak types signal θ_{NF} and strong types signal θ_F. Entrants choose E if they see θ_{NF} and ~E if they see θ_F. Now consider whether a defender can make a profitable deviation. Given that entrants will choose E if they hear θ_{NF}, a profitable deviation for a weak defender is to pool with the strong defenders and say θ_F (i.e., 300>160).

No separating equilibria exist where defenders signal by type. The case where defenders separate but signal their opposite type is clearly not an equilibrium.

Pooling

Proposition 4 *A pooling equilibrium exists where all defenders send θ_F, entrants enter if they hear θ_F and have off-path beliefs such that $P(\theta_{NF}) = 0$ and thus enter if they hear θ_{NF}, weak defenders backdown and strong defenders fight.*

Proof: In a pooling equilibrium no defender will have a strictly positive incentive to send the opposite signal. Consider that all defenders say θ_F. Now the signal has no information content and entrants will play based on their prior, which has not been updated by tht signal. In this case, they will always enter. Now consider whether, given certain entry, a weak type will prefer to signal θ_{NF}. If θ_{NF} is signaled they face certain entry and will choose to not fight, giving them 160. If there is pooling then the entrant does what they would do sans signal (enter), because the signal is uninformative.But then the weak defender also gets a payoff of 160 after signalling θ_F, and hence are indifferent between θ_F, θ_{NF}. Hence weak types do not have a positive incentive to change their signal. Strong types do not have an incentive deviate and send θ_{NF} either, as this will lead to entry (given the off path beliefs) and $300 > 160$. A similar logic applies to the indifference of a strong type from signalling θ_F or θ_{NF}. It can also be shown that there does not exist pooling where both defenders signal θ_{NF} because strong defenders will want to separate.

Note that the restriction on off path beliefs is what moves the analysis away from babbling where the defender places positive probability on all messages. This seems reasonable given the nature of the messages, where subjects could clearly map a relationship between the message, and an action to be played (given that the message said something about an action to be played), and the corresponding type that would play such an action.[1]

Semi-separating

Proposition 5 *A semi-separating equilibrium does not exist.*

[1] For a discussion of cheap talk refinements relevant here, see Joseph Farrell, 1993, Meaning and Credibility in Cheap-Talk Games, *Games and Economic Behavior*, 5, 514-531

Proof: Assume a semi-separating equilibrium exists. Under a semi-separating equilibrium the weak defender must be indifferent between signalling θ_F and θ_{NF}, which will arise because the entrant is mixing with probability $q < 1$ after observing θ_F. Now ask whether weak defenders can be made indifferent between signalling θ_F or θ_{NF}. As long as entrants do not enter with some probability, strong types will always send θ_F. In this equilibrium entrants will be mixing and entering with probability q if they hear θ_F and choosing E for sure if they hear θ_{NF}. Consider the strong type's EU from sending θ_F, $EU(\theta_F) = 300(1-q) + 160q = 300 - 140q$. Now consider their utility for sending θ_{NF}, $EU(\theta_{NF}) = 160$. For any $q < 1$ the strong type will strictly prefer to send the fight signal. Can entrants enter with some probability q after hearing θ_F such that they make weak defenders indifferent between θ_F or θ_{NF}? I.e., consider that the entrant selects E with probability q after observing θ_F, and select E with probability 1 after observing the no fight signal. Hence the expected utility of a weak type signalling θ_F is $(1-q)300 + 160q = 300 - 140q$. Their utility from signalling not fight is 160, because they will face entry for sure. Can they be made indifferent? They can only be made indifferent if $q = 1$. But then the entrant is no longer mixing and hence a semi-separating equilibrium does not exist. Is it possible to slightly modify the payoffs in order to have a semi-separating equilibrium?

Proposition 6 *There exists alternative payoff parameters for the game such that a a semi-separating equilibrium exists.*

Proof: Let x be the weak defender's payoff to not fighting after entry. Then they are indifferent between signalling θ_F and facing entry with probability q and signalling θ_{NF} and facing entry for sure if:

$$\begin{aligned}
(1-q)300 + qx &= 160 \\
300 - 300q + qx &= 160 \\
140 - q(300 - x) &= 0 \\
140 &= q(300 - x) \\
\frac{140}{300 - x} &= q
\end{aligned}$$

$q = 1$ when $x = 160$. q must be between 0 and 1. By definition of being a weak defender, x cannot be less than 70, and hence the range of values of x such that there could be a semi-separating equilibrium is $(70, 160)$.

Now imagine that $x = 100$. Then $q = \frac{140}{300-100} = .7$. Is there a semi-separating equilibrium? If $q < 1$ then strong types strictly prefer to send θ_F. Weak types still play ˜F after signalling either θ_F,θ_{NF}.

Now consider that the weak types signal fight with probability z. What are the entrants beliefs?

$$\begin{aligned}
\Pr(Weak|\theta_F) &= \frac{\Pr(\theta_F|Weak)\Pr(Weak)}{\Pr(\theta_F|Weak)\Pr(Weak)+\Pr(\theta_F|\tilde{\ }Weak)\Pr(\tilde{\ }Weak)} \\
&= \frac{\Pr(\theta_F|Weak)(1-p)}{\Pr(\theta_F|Weak)(1-p)+\Pr(\theta_F|\tilde{\ }Weak)p} \\
&= \frac{z-pz}{z-pz+p} \\
&= \frac{\frac{2}{3}z}{z-\frac{2}{3}z+\frac{1}{3}} = k
\end{aligned}$$

Hence in such an equilibrium the entrant's expected utility for entry following θ_F is $150k+(1-k)80 = 70k+80$. Their utility for not entering is 95 and hence are indifferent when

$$\begin{aligned}
70k+80 &= 95 \\
k &= \frac{15}{70}
\end{aligned}$$

For what probability, z, of the weak types signalling θ_F, could this hold?

$$\begin{aligned}
\frac{\frac{2}{3}z}{z-\frac{2}{3}z+\frac{1}{3}} &= \frac{15}{70} \\
\frac{2z}{z+1} &= \frac{15}{70} \\
z &= \frac{3}{25}
\end{aligned}$$

Hence under this alternative paramterization of the game there exists a semi-separating equilibrium where strong types always signal θ_F and play F if entered upon, weak types signal θ_F with probability $z = \frac{3}{25}$ and choose $\tilde{\ }F$ if entered upon, and entrants enter with probability 1 if observing θ_{NF} and enter with probability $q = .7$ after observing θ_F.

2.2 Repeated Play

What will happen in the repeated play version of the entry-deterrance game but where the defender can signal to each entrant they face? We use the notation from the first section, with the introduction of the message space $\Theta = (\theta_F, \theta_{NF})$. *Only signals in the present period are seen*–signals of the defender to earlier entrants are not seen. Hence we do not index signals by period. Again we index periods by labeling the last period 1, the second to last period 2, etc. The repeated play case with signalling is obviously more challenging.

As in the repeated play model with no signalling, if an entrant sees that the defender has previously not fought then their belief that they face a strong type is updated to 0 and they will enter for sure.

As before there will exist an entry constraint in terms of the entrant's beliefs, P, similar to $P_n < Q^n$. However, unlike before P_n may be influenced by both the history of the game but also the signal in the current period. Thus we have $P_n(\theta) < W$, where W does not necessarily equal Q^n.

Pooling equilibrium

Proposition 7 *There exists a pooling equilibrium where weak and strong defenders signal θ_F and all players play choose actions according to the sequential equilibrium of the repeated game without signalling.*

We must specify off path beliefs within any period. Here $P_n(\theta_{NF}) = 0$ is a reasonable assumption. Note, however, that these beliefs do not need to be transmitted as only action histories are recorded ($P_n(\theta_{NF}) \neq P_{n-1}$ unless the defender backs down in period n).

Proof: Note that signals in the final period will be uninformative and from the repeat play section without signalling we have posterior beliefs in the final period of $P_1(\theta_F) = P_1 = \frac{P_2}{P_2 + T_2(1-P_2)}$ (T_2 is the mixed strategy played by the defender in the second to last period). We will have the same final period constraint on entry (enter if $P_1(\theta_F) < Q = \frac{55}{70}$). In equilibrium the entrant still chooses E with probability $O_1 = .357$ (this is pinned down in the sequential equilibrium in order to make the defender in the second to last period indifferent), and weak defenders back down if entered upon.

Now ask whether a weak type in the last period would prefer to deviate to θ_{NF}. If they signal θ_{NF} the off path beliefs are such that they face guaranteed entry, $O_1 = 1$. Their expected utility in the pooling equilibrium is $300(1-O_1) + 160O_1$. This is greater 160 (what they would get if they faced entry for sure given the off path beliefs specified in equilibrium) for any value of O_1 less than 1. In equilibrium $O_1 = .357 < 1$. In the last period they still prefer to pool and signal θ_F.

Consider now a deviation that has defender separate in the second to last period and pooling in the final period. Will the weak defender want to deviate and signal θ_{NF} in the second to last period? Assume they deviate and signal θ_{NF} in the second to last period (but continue to pool in the last period). They face entry for sure in the second period. But if they end up fighting their reputation is sustained and they can still expect to face the mixture O_1 in the final period. Hence this deviation still leads them to the expected utility calculation of $T_2^*(70+160O_1^*+300(1-O_1^*))+(1-T_2^*)(160+160)$. This is exactly what they can expect to get if pooling and hence there is no positive incentive to deviate. If, alternatively, they decide to not fight in the second to last period they are clearly worse off than remaining in the pooling equilibrium, as in this case they lose their reputation and now face certain entry in the next period. Checking for profitable deviations in previous periods yields similar results.

Separating equilibrium

Proposition 8 *There does not exist a separating equilibrium where defenders pool until the last period and then signals by type.*

Proof: Assume a separating equilibrium exists in the final period where defenders signal by type, weak defenders back down if entered upon and strong types fight, entrants choose ˜E if they hear θ_F (with beliefs $P_1(\theta_F) = 1$) and E if they hear θ_{NF} (with beliefs $P_1(\theta_{NF}) = 0$). I.e., we would have $\Pr(Strong|\theta_F) = 1, \Pr(Fight|\theta_F) = 1, \Pr(Fight|Strong \cap \theta_F) = 1$, and $\Pr(Strong|\theta_{NF}) = 0$. Hence $P_1(\theta_F) = 1$. In this case there are no off path beliefs to specify because there are no probability 0 signals.

Now consider whether the weak type would prefer to signal θ_F. If $P_1(\theta_F) = 1$ then the entry constraint cannot be satisfied nor is it incentive compatible for the strong defender to play a mixed strategy to try and make the entrant indifferent. The entrant will not enter after hearing θ_F. Hence the weak defender would strictly prefer to send θ_F because 300 (payoff to defender from no entry) is greater than 160, which is what they would get if they faced entry. Hence there is no separating equilibrium in the final period.

Proposition 9 *There exists no separating equilibrium that begins with separation in the second period.*

There are two possibilities depending on whether separation also occurs in the final period. 1) Separation in the second to last period and the last period or 2) separation in the second to last period but pooling in the last period (recall signals are not transmitted to the next entrant).

1) Assume an equilibrium exists with separation by type in the last two periods. Here the entrant in the second to last period chooses E if θ_{NF} (they have beliefs $P_2(\theta_{NF} = 0)$ and ˜E if θ_F (they have beliefs $P_2(\theta_F = 1)$). A defender who sends θ_{NF} in the second to last period will face entry for sure. In this equilibrium after sending θ_{NF} the weak defender gets either 160 or 70 (depending on whether they choose to fight). In the final period of this equilibrium they also choose to send θ_{NF}, face certain entry, and do not fight (there is no value of the reputation). This gives them 160.

Consider now a deviation where the weak type sends θ_F in the second to last period. Then they will get 300 plus some amount from the final period ($P_2(\theta_F) = 1$ and so the entrant does not enter). However, no matter what outcome happens in the final period the sum of their utility from signalling θ_F

in the second period will always be greater than the most they could get by signalling by type beginning in the second to last period: $160 + 160 = 320$. Hence there will be no seperation by type beginning in the second to last period and continued in the last period. A similar logic shows that there will be no *sustained* separation starting in any period.

2) Now consider an equilibrium with separation in the second to last period but pooling in the last period. The appropriate deviation to check is the single shot deviation to pooling in the second period, but maintaining the pooling in the final period. As described above the weak defender has no incentive to deviate and play the separating strategy in the final period, but we still need use a single shot deviation to check the posited equilibrium. In the second to last period of this potential separating equilibrium the weak defender will face entry for sure $P_2(\theta_{NF}) = 0$ and will get either 70 or 160, depending on whether they fight. With $P_2(\theta_{NF}) = 0$ the entrant enters with probability $O_2 = 1$. Then the payoff to the entrant is $T_2^*([70]+[160O_1^*+300(1-O_1^*)])+(1-T_2^*)([160]+[160]) = T_2^*[(370-320)-O_1^*(300-160)]$, where T_2^* is the mixture played by the weak defender in the second to last period after signalling θ_{NF}, the terms inside the [] are the period 2 and period 1 payoffs following the realization of T_2^*, and O_1^* is the mixture played in the last period played by the entrant.

Now consider the deviation where they pool in the second to last period. In this case $P_2(\theta_F) = 1$, the entry threshold for beliefs does not hold and the entrant does not enter. The defender gets a payoff of 300 and then faces the next entrant, yielding an expexted utility of $300 + 160O_1^{**} + 300(1 - O_1^{**}) = 600 - 140O_1^{**}$. Here O_1^{**} is the potentially mixed strategy played by the final entrant following no entry by the previous entrant.[2] Under the posited equilibrium they would get $T_2^*[(370-320)-O_1^*(300-160)]$. The deviation is profitable if:

$$\begin{aligned} 600 - 140O_1^{**} &> T_2^*((370-320)-O_1^*(300-160)) \\ 600 - 140O_1^{**} &> T_2^*(50-140O_1^*) \end{aligned}$$

For all values of O_1^{**} and O_1^* this inequality will hold and hence there exists a profitable deviation from the posited equilibrium to pool in the second to last period. Furthermore, the maximum size of the RHS is when $O_1^* = 0$ and $T_2^* = 1$,

i.e., a value of 50. Alternatively, the deviation yields 460 when $O_1^{**} = 1$.

The next step would be to repeat the process and show that there would be no separation in the third period. The opportunity to avoid sure entry is preferable to the defender, but it seem clear that this is unlikely to occur given that within period signals of θ_{NF} lead to updated beliefs of $P_n(\theta_{NF}) = 0$.

Semi-separating equilibrium

A final type of equilibrium to check is whether semi-separation occurs.

[2] The section on repeated play without communication shows that in this case the entrant enters for sure because their belief threshold is exceeded. I abstract from this so as to illustrate the logic but it clearly does not matter.

Proposition 10 *There is no equilibrium where pooling is played until the final period and then semi-separation occurs. Assume that $P_1(\theta_{NF}) = 0$ since strong types will never signal θ_{NF}.*

Proof: Assume there is a semi-separating equilibrium in the last period. In this period the entrant must play a mixed strategy, O_1, that satisfies the indifference condition for the weak defender in the second to last period, $T_2[70 + 160O_1 + 300(1 - O_1)] + [1 - T_2](320) = 320$, but also makes weak defenders indifferent between signalling θ_F or θ_{NF}. Let the weak defender signal θ_F with probability z_1. To solve for this mixture we need to find the value of O_1 that makes the defender indifferent about whether or not to signal fight. Furthermore, this value of O_1 must also make the defender in the second to last period indifferent between fighting or not in order to sustain the mixed strategy equilibrium in actions. Finally, the the influence of z_1 and T_2 on the entrants beliefs must also make the entrant want to mix (and hence be indifferent between entering or not) after observing θ_F.

The expression for O_1 will contain T_2 and the expression for T_2 will contain $P_1(\theta_F)$. Assume that up through the second to last period the reputation has been maintained, and hence the posterior following this period is P_2 (which forms prior for period 1). Given this posterior, the entrant in the last period observes a signal and can further update their beliefs about the type of defender they face.

Consider first the entrant's beliefs in the final period (though the contradiction will be obtained from showing the defender cannot be made indifferent).

In order to check for beliefs under which the entrant can be made indifferent about entry in the final period, we need their beliefs, P_1, which is their posterior from observing the previous history of play and, with communication, the signal from the defender. Given the restriction on beliefs that if backing down occurs this prior is updated to 0, the entrant's belief in the final period is $\Pr(Strong|\Theta, F_2)$. Consider the beliefs following signalling θ_F (in semi-separation strong types never signal θ_{NF} and hence $\Pr(Strong|\theta_{NF}, F_2) = 0$).

By Bayes Rule:

$$
\begin{aligned}
\Pr(Strong|\theta_F \cap F_2) &= \frac{\Pr(\theta_F \cap F_2|Strong)\Pr(Strong)}{\Pr(\theta_F \cap F_2)} \\
&= \frac{\Pr(\theta_F \cap F_2|Strong)\Pr(Strong)}{\Pr(\theta_F \cap F_2|Strong)\Pr(Strong) + \Pr(\theta_F \cap F_2|\tilde{\ }Strong)\Pr(\tilde{\ }Strong)} \\
&= \frac{1P_2}{1P_2 + \Pr(\theta_F \cap F_2|\tilde{\ }Strong)(1-P_2)} \\
&= \frac{P_2}{P_2 + \Pr(\theta_F|\tilde{\ }Strong)\Pr(F_2|\tilde{\ }Strong)(1-P_2)} \\
&= \frac{P_2}{P_2 + T_2 z_1(1-P_2)} \\
&= \frac{P_2}{P_2 + T_2 z_1 - P_2 T_2 z_1}
\end{aligned}
$$

The entrant's expected utility calculation for choosing E following signal θ_F is $150(1-P_1(\theta_F)) + 80(P_1(\theta_F)) = 150 - 70P_1(\theta_F)$. Hence they are indifferent only if $150 - 70P_1(\theta_F) = 95$, or $P_1(\theta_F) = \frac{55}{70} = \frac{P_2}{P_2+T_2z_1-P_2T_2z_1}$.

Solving for z_1 we have:

$$
\begin{aligned}
\frac{55}{70} &= \frac{P_2}{P_2 + T_2 z_1 - P_2 T_2 z_1} \\
\frac{55}{70}(P_2 + T_2 z_1 - P_2 T_2 z_1) &= P_2 \\
\frac{55}{70}T_2 z_1 - \frac{55}{70}P_2 T_2 z_1 &= P_2 - \frac{55}{70}P_2 \\
z_1(\frac{55}{70}T_2 - \frac{55}{70}P_2 T_2) &= P_2 - \frac{55}{70}P_2 \\
z_1 &= \frac{P_2 - \frac{55}{70}P_2}{\frac{55}{70}T_2 - \frac{55}{70}P_2 T_2}
\end{aligned}
$$

Notice how the exact mixture played by the defender depends on the entrant's prior belief (from the second period) and the mixture T_2 played by the defender in the previous period. In the model without communication, we simply solved for T_2 in order to establish the indifference condition. Here, this would be:

$$
\begin{aligned}
\frac{55}{70} &= \frac{P_2}{P_2 + T_2 z_1 - P_2 T_2 z_1} \\
\frac{55}{70}T_2 z_1 - \frac{55}{70}P_2 T_2 z_1 &= P_2 - \frac{55}{70}P_2 \\
T_2 &= \frac{P_2 - \frac{55}{70}P_2}{\frac{55}{70}z_1 - \frac{55}{70}P_2 z_1}
\end{aligned}
$$

The entrant wants to choose a mixture O_1^* to make the defender in the second to last period indifferent between fighting in the second to last period AND indifferent between signalling θ_F or θ_{NF}. Both know that the weak type will play a mixture z_1 in the final period if fight is chosen in the second to last period. Consider the first indifference condition:

$$\begin{aligned}
T_2^*(70 + (1 - z_1)160 + 300z_1(1 - O_1^*)) + (1 - T_2^*)(160 + 160) &= 320 \\
T_2^*(70 + 160 - 160z_1 + 300z_1 - 300z_1O_1^*) + 320 - 320T_2^* &= 320 \\
140z_1T_2^* - 90T_2^* - 300z_1O_1^*T_2^* + 320 &= 320 \\
140z_1T_2^* - 90T_2^* - 300z_1O_1^*T_2^* &= 0 \\
-300z_1O_1^*T_2^* &= -140z_1T_2^* + 90T_2^* \\
O_1^* &= \frac{-140z_1T_2^* + 90T_2^*}{-300z_1T_2^*} \\
O_1^* &= \frac{140z_1 - 90}{300z_1}
\end{aligned}$$

Notice that z_1 must be greater that $\frac{90}{140}$, otherwise the expression is negative.

Finally, it must also be the case that the weak defender is indifferent between signalling θ_F or θ_{NF} in the final period, given some mixture by the entrant O_1^*. The expected utlity of sending fight is $160O_1^* + 300(1 - O_1^*) = 300 - 140O_1^*$. The expected utility of sending θ_{NF} is 160 because the entrant will enter for sure.

$$300 - 140O_1^* = 160$$

Hence this indifference condition can only be satisfied if $O_1^* = 1$. But in such a case the entrant is no longer indifferent between entry. Furthermore, this pins down the previous indifference condition to be

$$\begin{aligned}
O_1^* &= \frac{140z_1 - 90}{300z_1} \\
1 &= \frac{140z_1 - 90}{300z_1} \\
160z_1 &= -90 \\
z_1 &= \frac{-90}{160}
\end{aligned}$$

Which cannot happen. There is no semi-separating equilibrium that begins in the final period of the game.

Proposition 11 *There is no equilibrium where pooling is played until the second to last period, in period 2 semi-separation occurs, and in the final period weak defenders pool. Again assume that $P_2(\theta_{NF}) = 0$ since strong types will never signal θ_{NF}.*

Proof. Assume an equilibrium exists with semi-separation in the second to last period and pooling in the final period. For such an equilibrium to exist there must be a O_2 such that the weak defender is indifferent between signalling θ_{NF} or θ_F. Within the second period a signal of θ_{NF} will lead to sure entry whereas θ_F will be followed by some mixture O_2. O_2 enters into the defender's utility function in the third period. Hence we have the expected utility of sending θ_F.

$$
\begin{aligned}
& T_3(70 + T_2^*(70O_2 + 300(1 - O_2) + 160O_1^* + 300(1 - O_1^*)) + (1 - T_2^*)(160O_2 + 300(1 - O_2) + 160)) \\
& +(1 - T_3)(160 + 160 + 160) \\
= \; & 50T_3 - 140O_2T_3 + 140T_3T_2^* - 90O_2T_3T_2^* - 140T_3O_1^*T_2^* + 480
\end{aligned}
$$

Similarly, if the signal in the second period is θ_{NF} then we have

$$
\begin{aligned}
& T_3(70 + T_2^*(70 + 160O_1^* + 300(1 - O_1^*)) + (1 - T_2^*)(160 + 160)) + (1 - T_3)(480) \\
= \; & 50T_3T_2^* - 90T_3 - 140O_1^*T_3T_2^* + 480
\end{aligned}
$$

Hence indifference implies:

$$
\begin{aligned}
50T_3 - 140O_2T_3 + 140T_3T_2^* - 90O_2T_3T_2^* - 140T_3O_1^*T_2^* + 480 &= 50T_3T_2^* - 90T_3 - 140O_1^*T_3T_2^* + 480 \\
50T_3 - 140O_2T_3 + 90T_3T_2^* - 90O_2T_3T_2^* &= -90T_3 \\
140 - 140O_2 + 90T_2^* - 90O_2T_2^* &= 0 \\
140 + 90T_2^* - 140O_2 - 90O_2T_2^* &= 0 \\
140 + 90T_2^* &= 140O_2 + 90O_2T_2^*
\end{aligned}
$$

But the only way this can happen is if $O_2 = 1$. But then the defender can't be mixing their signal in a way that makes the entrant indifferent, and the entrant can't be mixing their strategy to make the defender in the previous period indifferent. Semi-separation does not happen in the second period.

A similar logic applies to previous periods.