In "Reputation & Civil War" by Barbara F Walter
Cambridge UP

3 An experimental study of reputation building and deterrence

(co-authored with Dustin Tingley)

This chapter uses a series of laboratory experiments to test whether individuals react to incentives to build reputations in the way the chain store model predicts.[1] We include this chapter because experiments offer a unique opportunity to adjust or improve the theory before it is tested against the complex and often messy data on self-determination disputes, and because it offers an unambiguous way to determine causality. If subjects adhere to the predictions of the theory in the controlled environment of the laboratory, then we know that people are at least capable of making the strategic decisions outlined in Chapter 2. If subjects deviate from the predictions, then we know there could be a gap between the theory and behavior that should be flagged before testing the theory in the real world.

Still, the strength of laboratory experiments lies in their ability to test the logic of a theory, and not their applicability to more real-world settings. Showing that undergraduate students at Princeton begin to build reputations in response to strict incentives offered in the laboratory does not mean that government leaders will do the same in the more complex arena of domestic politics. Determining the relevance of the theory to actual disputes between governments and minority groups requires a different set of tests, which will be presented in Chapters 4 through 7.[2] Because different empirical approaches all have significant weaknesses, our research design

[1] For an excellent overview of experimental methods in political science, including their strengths and weaknesses, see McDermott (2002).
[2] See Morton and Williams (2007) and Shadish *et al.* (2002).

draws on multiple methodologies and tries to triangulate across these approaches.[3]

A final reason we include this chapter is that two of the more recent and prominent criticisms of reputational theory in international relations draw from laboratory experiments to help buttress their claims. Both Mercer, and to a lesser extent Press, claim that experiments in psychology laboratories give credence to their claims. But the experimental studies they cite were not designed for testing the type of reputational arguments we have in mind. Moreover, a number of *incentivized* experiments they do not cite suggest reputation building does matter. To maintain a consistent methodological approach with these more recent studies, we designed an experiment that focuses on the key elements of reputational theory outlined in the previous chapter.

THE EXPERIMENT AND ITS OBJECTIVES

This chapter attempts to assess two important aspects of the chain store model.[4] The first is whether the past behavior of the defender affects decisions to challenge, and if so in what ways. According to the model, entrants should be far more likely to challenge if they observe even a single act of accommodation, and these challenges should begin immediately after accommodation is observed. The laboratory experiments, therefore, allow us to determine whether this temporal relationship exists.

This chapter also seeks to determine whether the number of future challengers can cause defenders and entrants to change their behavior. We know from the model that repeated play creates

[3] See Kinder and Palfrey (1993), McDermott (2002), and Bueno de Mesquita (2002).

[4] We are not the first scholars to design experiments to test the chain store model. A number of experiments have tested different implications of the model. See Bolton and Ockenfels (2007), Brandts and Figueras (2003), Camerer and Weigelt (1988), Jung *et al.* (1994), Neral and Ochs (1992), and Sundali *et al.* (2000). We are the first to explore experimentally a particular comparative static of the equilibrium model – changing the number of entrants a defender faces.

incentives to build reputation, but we do not know how many players are needed before this occurs.[5] We also do not know whether players significantly alter their behavior based on the number of iterations they expect to encounter.

The experiment reveals that subjects do engage in reputation building behavior in a repeated game with incomplete information, and that reputation building deters future challenges under these conditions. As will be discussed in greater detail below, conciliatory defenders do choose to fight against early challengers in order to signal toughness even though they could potentially pay heavier short-term costs as a result. Similarly, entrants are significantly more likely to challenge after they have observed accommodation. Thus, in a controlled laboratory environment where the amount of information and the number of repeated plays can be controlled, reputation building emerged as an important variable influencing behavior.

But the experiments also suggested two ways in which behavior deviated from the expectations of the model. The first has to do with the number of repetitions necessary for reputation building to emerge. Subjects invested in reputation building even against a relatively small number of challengers. Thus, depending on the payoffs, a game need not be repeated as often as theory might predict for reputation building to emerge. The second deviation has to do with entrant and defender behavior in the very early rounds of the game with a large number of entrants. Entrants challenged much sooner, and defenders accommodated these early entrants much more than the theory predicted. These results suggest that at least three additional hypotheses regarding government and ethnic group behavior need to be added. We discuss these findings in more detail below.

[5] One exception is Richard Selten (1978) who suggested that reputation building was unlikely to occur in games 1–4, would begin to break down around games 5–8, and would certainly not occur at game 20, but these critical thresholds have never been empirically tested.

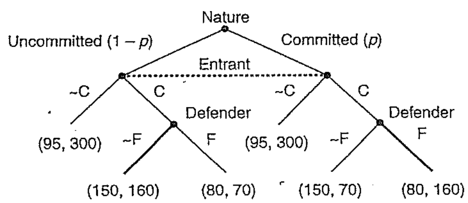

FIGURE 3.1 The structure of a single-shot play

THE STRUCTURE OF THE GAME

Figure 3.1 shows the structure of a single-shot play of the game we were attempting to test as well as the payoffs each of the players knows it will receive for the different outcomes.[6]

The game is the same repeated version of the entry–deterrence game with incomplete information used in Chapter 2.[7] The game begins with nature randomly choosing whether the defender is committed or uncommitted to fighting a challenge. If the defender is committed, the entrant should prefer not to challenge. If the defender is uncommitted, the entrant should prefer to fight. The entrant, however, does not know what type of defender it faces. If the entrant decides to challenge, the defender then chooses whether to fight this challenger or make the concessions that are necessary to avoid a fight. Once the defender chooses whether to fight or accommodate, a second entrant then chooses whether to challenge, after which the defender again decides whether to fight or accommodate. As

[6] Note that the Defender's (D) payoffs depend on his or her type. If the entrant chooses to challenge, a committed defender receives a higher payoff from fighting, while an uncommitted defender receives a higher payoff from backing down. Payoff parameters are the same as those used by Jung *et al.* (1994).

[7] The standard set-up of this game – which gives rise to the so called "chain store paradox" – assumes one-sided incomplete information because the entrants in the model (firms) are assumed to be homogeneous. This is not the case in many situations in the real world. Some entrants may put up more of a fight than others, and defenders may not have perfect information about the "type" they face. Future research is needed to incorporate two-sided incomplete information into experiments (see Tingley and Wang 2008), although we note that working out tight theoretical predictions in such cases with repeated play would be quite difficult.

each entrant plays, they are given information about how previous entrants played against the defender they are currently matched with, and the choice of the defender *if* the previous entrant decided to challenge.[8] The game continues until the defender has faced a commonly known set number of entrants.[9]

In the actual experiment, we used value neutral labels to avoid leading the subjects in any way. Instead of entrant, we used "first-mover," and instead of defender we used "second-mover." In each interaction with a potential challenger, the defender was first asked "if the first-mover chooses 'A1' (enter), what will be your response?" This question was asked because entrants were expected not to enter in early rounds, thus making it impossible to determine how the defender would respond to a challenge at this point. Including this question, therefore, allowed us to collect data both on and off the actual path of play. In the experimental literature this is known as the strategy solicitation method.[10]

In what follows, we explore the sequential equilibrium predictions of this repeated game and then return to a more thorough description of the experiment and our results.

[8] Challengers do not, however, know what payoffs a defender receives from each of his or her decisions. This information is not revealed because it would also reveal the state's type in cases where they challenge entry.

[9] 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 the ethnic group demanded greater and greater concessions over time. The Canadian government's relationship with the Parti Quebecois would be an example of this type of two-player repeated game. In order to simplify the analysis we only analyze repeated play between different opponents. The same dynamics, however, should hold in any situation where a defender expects repeated interactions over time.

[10] This does not mean that our subjects played the simultaneous normal form version of the game, as the defender's strategy choice is only elicited for the conditional case where there is entry. Our empirical results conform to previous experiments on the chain store game that did not use the strategy method, and thus our experiment *appears* to reside in the set of cases where the strategy solicitation method has little effect on strategy choices (Brandts and Charness 2000).

HYPOTHESES

The model presented above has a sequential equilibrium that is (plausibly) unique, meaning that actors are best responding at every sequential stage of the iteration, updating their beliefs with relevant observations. The sequential equilibrium prediction of this game can be summarized as follows.[11] If an entrant observes a defender accommodating one challenger, its belief that the defender is committed is updated to $p=0$. Knowing that the defender is not committed, all subsequent entrants should challenge, and all uncommitted defenders should continue to accommodate. This portion of the equilibrium generates the following hypotheses:

> $H1_E$: *entrants should always enter if they observe accommodation.*
>
> $H1_D$: *weak defenders that have backed down once should continue to back down.*

Not all entrants, however, will observe accommodation in a previous round. If the entrant observes no accommodation, we consider the game in three phases, an early, a middle and a late phase. This is where we should observe differences in behavior based on the number of challengers a defender anticipates. If the number of challengers is sufficiently high, there is a pure strategy period of play where entrants never enter and defenders always fight. This is because entrants understand that even weak defenders have incentives to invest in reputation in these early rounds. After reaching some period, φ^*, defenders have incentives to play a mixed strategy, fighting with some probability, and backing down with some

[11] The analysis provided here follows Jung *et al.* (1994) to which we refer the reader for the mechanics of the proof. We note that the sequential equilibrium also prescribes different strategies for entrants depending on whether there was entry by the previous entrant in the immediately preceding period. While technically an important part of the equilibrium, the results below do not change if we break apart the data to accommodate this fact. To simplify our presentation and tighten our focus on reputation building we ignore this feature.

probability. Entrants should respond accordingly. During the mixed strategy period, defenders should fight with a decreasing probability as the end of the round nears. On the last round of play, weak defenders should always back down and challengers should always enter.

This generates the following set of hypotheses in those cases where no previous accommodation was observed:

> $H2_E$: *if the period* $< \varphi^*$ *then entrants should enter with probability 0.*
>
> $H2_D$: *if the period* $< \varphi^*$ *then weak states should fight with probability 1.*

VARYING THE NUMBER OF FUTURE CHALLENGERS

We investigate two different experimental treatments that vary the total number of future challengers a defender faces. In one treatment, a defender faces eight future entrants, in the other the defender faces four. These numbers were chosen for the following reason. Previous research on self-determination movements found that governments invested in reputation building when faced with as few as four potential separatist challengers.[12] The theoretical model considered here, however, suggests that for the payoffs we use, reputation building should only consistently occur in cases with *more* than four challengers. Our four- and eight-entrant design was chosen to mimic this feature of the real world while also setting up a stark difference in the predictions of our model to test in the laboratory.

According to the model, the sequential equilibrium predictions about play in the four-entrant and eight-entrant cases are different in an important way. In the eight-entrant design there is an initial pure strategy phase described above: entrants should never challenge, and any challenge should always be resisted. The mixed-strategy

[12] See Walter (2006).

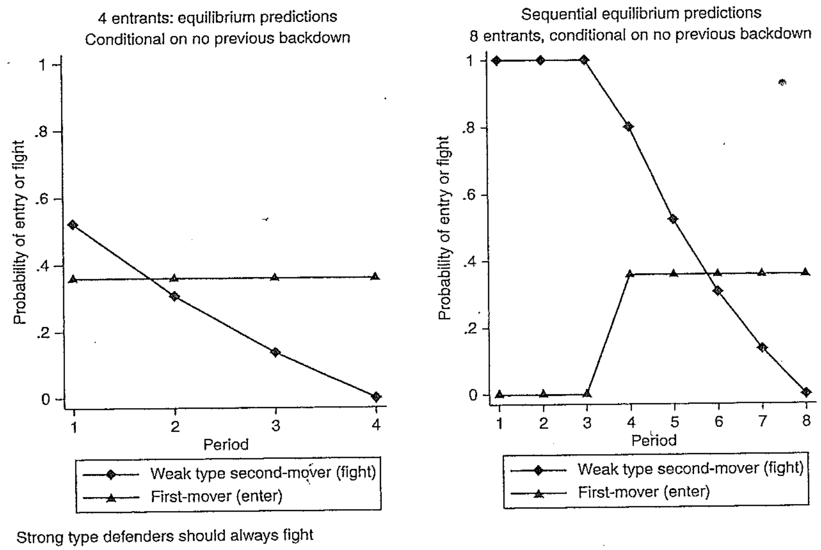

FIGURE 3.2 Predicted probability of entry and fighting in each period

phase of the game should not begin until the fourth round in the eight-entrant design.

When there are four entrants, subjects will be in the mixed strategy phase of play from the very beginning, meaning that much less reputation building should occur in early rounds. Entrants should be much more apt to challenge in the first round of play, and defenders should be more likely to back down. The behavior of the defenders and challengers in this design should mirror the behavior we observe in the last four rounds of the eight-person design.

We can graphically illustrate the equilibrium predictions for different numbers of entrants. Figure 3.2 shows the predicted probability of entry and fighting against each period of the sequence, *given that the defender has not backed down in the past.* Figure 3.2 does not include strong type defenders, as they should always fight in every period.

These graphs illustrate the fundamental difference between the four- and eight-entrant designs, and suggest the following testable hypotheses:

$H3_E$: *in the four-entrant case, we should see more entry in the first, second, and third periods of play compared to the eight-entrant case.*

$H3_D$: *in the four-entrant case, we should see more accommodation in the first, second, and third periods of play compared to the eight-entrant case.*

The intuition behind this result is that in the four-entrant treatment the final round of play is closer, creating fewer incentives for defenders to invest in reputation building in early rounds.

EXPERIMENTAL DESIGN

We recruited subjects through Princeton University's Laboratory for Experimental Social Science (PLESS) using an e-mail solicitation to all Princeton students who had signed up with the lab. Those who responded were accepted until all positions were filled. 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 the 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 was done to ensure that all aspects of the experiment design were common knowledge. A copy of the instructions is included in Appendix 1.

Subjects were randomly assigned to two separate groups, entrants and defendants, which were referred to simply as first movers and second movers. Second movers (defenders) were also assigned a "type" (weak or strong), which we called "type 1" or "type 2." First movers (entrants) were only told that there was some probability ($p = \frac{1}{3}$) that they were facing a strong type of defender. All of the subjects were told exactly how many total defenders or entrants they would face.[13]

[13] The fact that the terminal node was known to all subjects does not always exist in the real world. Defenders often do not know exactly how many interactions

Defenders faced a known number of entrants, and each entrant played each defender once within a "run" of the experiment.[14] Entrants were able to see the history of the defender they were paired with, which included how all previous entrants had played, and how the defendant had responded.

The experiment proceeded as follows. Entrants were asked to choose between A1 and A2 (entry and not entry). Defendants 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). In the next period entrants were matched with a new defender. Each entrant would observe what happened in this defender's earlier periods. As a result of this matching procedure, each entrant made one decision with no available history (in the first period), one decision with information on how their defender played in the first period with a different entrant, and so on. 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. Subjects were allowed to write these payoffs down to compare with future runs. Payoffs were expressed in points which were then translated to US dollars at the end of the experiment.

The experiment was repeated four times in order to take into account the effects of learning and to generate sufficient data for the analysis. The precise number of times the subjects would repeat the experiment was unknown to them. 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). Appendix 1 covers our design in more detail.[15]

they are likely to have. Governments, for example, might be able to reasonably estimate how many ethnic separatists they are likely to encounter, but they are unlikely to know when the last challenge will be. The fact that the final round of play is unknown, however, does not mean that reputation building will not occur. In fact, this uncertainty is likely to create greater incentives to invest in reputation building, since defenders know they have declining incentives to invest in reputation as the terminal node nears.

14 Subjects were informed of their roles at the beginning of a run; first and second movers kept their roles throughout the entire experiment, but the second mover "type" could change from one repetition of the experiment to another.

EMPIRICAL ANALYSIS

The model we test is about reputation building, and thus our analysis is crafted in a way that conditions behavior on the history of play by each defender and observed by each entrant. Our basic empirical strategy is to derive empirical probabilities of entry and fight behavior by taking mean values of the dichotomous decision variable *conditional on the history of play experienced by each subject*. Consider first the case where an entrant faces a defender that has not backed down in an earlier period. For each individual period (1–4 or 1–8) we add up the number of enter (or fight) decisions that occur where the defender had not previously backed down in an *earlier* period. We then divide by the total number of observations where the defender had not previously backed down in an earlier period. Because we code our Enter variable as a 1 if Enter (Fight) was chosen and 0 if Not Enter (Not Fight), this gives us the probability of Enter (Fight) conditional on the fact that the defender had not previously backed down in an earlier period. We calculate these values for each period. We perform an analogous calculation for cases where the defender had backed down in an earlier period.[16]

15 Note that our design, instructions, and computer interface went through a lengthy series of reformulations/testing in order to obtain the best possible experimental protocol.

16 This is the appropriate empirical strategy given that the strategic situation has players using mixed strategies. We do not directly observe the randomization procedure that subjects (presumably) utilize, and thus by aggregating across subjects we assume that each individual uses the same mental randomization device. This is a standard empirical approach when analyzing laboratory (O'Neill 1987) and field data (Walker and Wooders 2001), but also represents an active area of debate in the experimental literature (Ochs 1995). We consider a more individual level analysis below.

RESULTS AND INTERPRETATION

The effects of past accommodation

The results are organized into two sections. First, we look at whether entrants changed their behavior based on the past behavior of defenders. According to the sequential equilibrium prediction, entrants should always enter if they observe a defender backing down in a previous period. This gives us a direct check on whether a failure to be tough makes challenges more likely.[17]

Figure 3.3 shows the average behavior of both the entrant and the defender in cases where the defender backed down in a previous period. What we see is striking. In line with predictions of the model, entrants in both the four- and eight-subject experiments almost always challenged if they had observed prior accommodation. This suggests that challengers updated their beliefs about what type of defender they were facing, almost certainly believing they were facing a weak defender if he or she had acquiesced. In addition, defenders in both experiments consistently chose to continue acquiescing once accommodation had occurred. This is strong evidence that entrants are strongly influenced by the past behavior of the defender. It also reveals that defenders generally do not attempt to rebuild a reputation for toughness once they have lost it.

Figure 3.3, however, reveals some behaviors that warrant closer examination. A small percentage of entrants in both the four- and eight-person experiments chose not to enter after the defender had revealed itself to be weak. In fact, as you can see on the left side of Figure 3.3, the probability of entry declined slightly in both cases after an initially high rate of entry. Because this behavior was quite rare, however, we take the overall results to be very supportive of $H1_E$ and $H1_D$.

[17] Each separate experiment session drew subjects from the same population, although subjects that had participated in our experiment previously were not allowed to participate again. Thus we pool across different sessions to generate separate samples for our eight and four entrant designs.

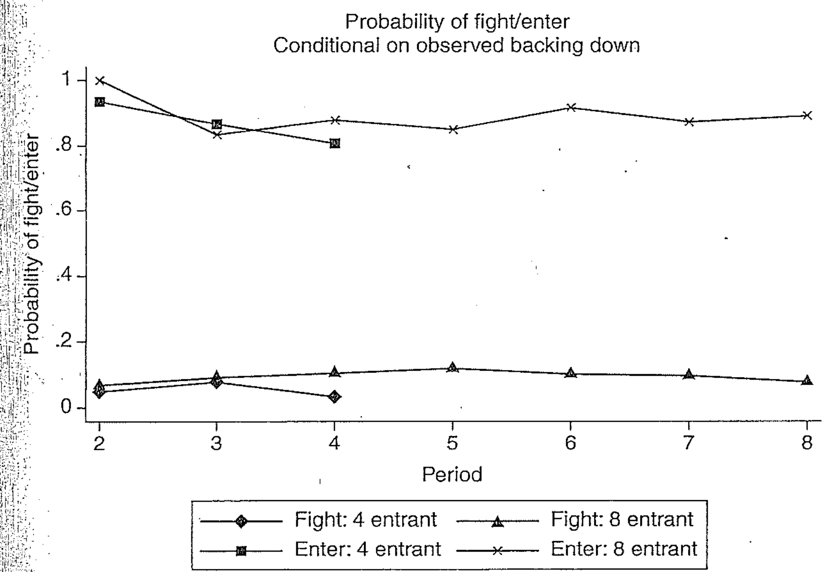

FIGURE 3.3 Probability of fight/enter

The effects of repeated play

Next, we looked to see if subjects behaved differently depending on the number of potential challengers that existed. Here we analyze per period probabilities of entry and fighting in cases where no accommodation was observed, aggregated over four separate "repetitions" of the experiment from each experimental session.[18] Figure 3.4 gives us our first look at whether the comparative static prediction on number of entrants holds in the laboratory.

Figure 3.4 reveals the first real deviation from the model. The model predicted that we should see significantly lower rates of entry in the first two periods in both the four- and eight-person designs than we observed in the laboratory. According to the theory, approximately 36 percent of entrants in the four-entrant design should

[18] Such an analysis ignores subject learning but is the appropriate first swipe at analyzing the data.

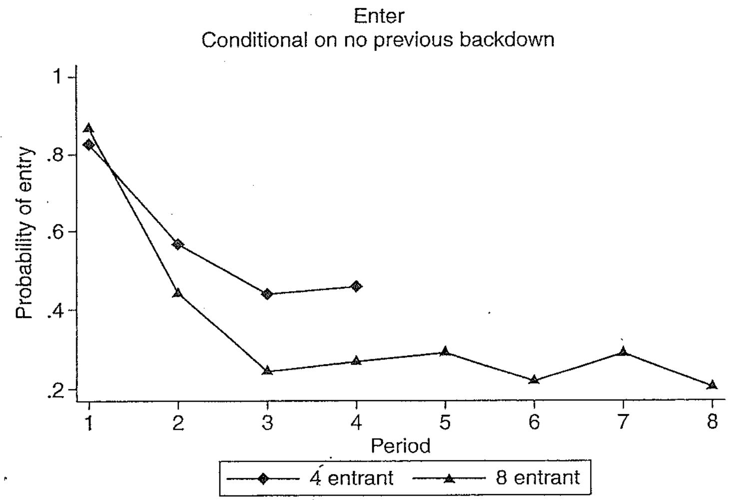

FIGURE 3.4 Probability of enter

enter in each period if there is no observed accommodation. Yet, as Figure 3.4 shows, over 80 percent of subjects in the four-entrant game entered in the very first round. This percentage dropped in the second and third round, but never reached the low rate predicted by the model.

The same is true in the longer eight-entrant game. According to the model, entrants should never enter in the first, second, and third periods of an eight-entrant design. After that, approximately 36 percent of them should enter if no previous accommodation is observed. But as Figure 3.4 shows, challengers again entered at a high rate in the first period, reduced their entry in the second period, and then ended close to the predicted entry rate of .357 by the third period. Thus, while the entry rate is higher than predicted in the first two periods of play, subjects behaved closer to equilibrium predictions during the mixed and final phases of the game.

In both designs, the striking differences between predicted and observed behavior occur in the very earliest rounds. The first three entrants in the eight-entrant design simply *should not enter.* If entrants enter during these rounds, and defenders fight them (as

the theory expects them to do), entrants will receive fewer payoffs than if they had not entered. Yet, over 80 percent of subjects in both designs chose to challenge in the very first round.

This behavior is quite puzzling and does not lend direct support to hypothesis $H2_E$. There are at least three possible explanations for this pattern. The first is that early entrants have not yet learned how to play the game, an explanation we explore in greater detail below. A second possibility is that entrants at the front of the pack challenge in order to gather information about defender type that could be valuable to the remaining entrants. This altruistic behavior is not taken into consideration by the model.[19] A third explanation is that entrants bring certain "homemade priors" or beliefs to the game, where they expect certain defenders to concede, making early entry potentially more profitable than the theory predicts.[20] A final explanation, which we investigate below, is that defender responses made these higher rates of entry rational.

We find mixed support for $H3_E$ which predicted that in periods 1–3 there should be more entry in the four-entrant design than in the eight-entrant design. In the first period there was no discernible difference in entry behavior. However, in periods two and three entry was significantly more likely in the four-entrant design.[21] The significant difference in the fourth period is not predicted by our theory, as the probability of entry in both designs should be .357. Our model predicts that in periods one, two, and three, the conditional entry rate should be higher in the four-player design. Thus, in two out of the three periods we find behavior that is consistent with $H4_E$.

Figure 3.5 shifts the focus onto defenders. Our equilibrium model predicted that defenders would fight based in part on whether

19 For a discussion of considerations along these lines see Bolton and Ockenfels (2007).

20 Camerer and Weigelt (1988).

21 Difference in means and difference in proportions both produce significant test statistics with $p<.05$.

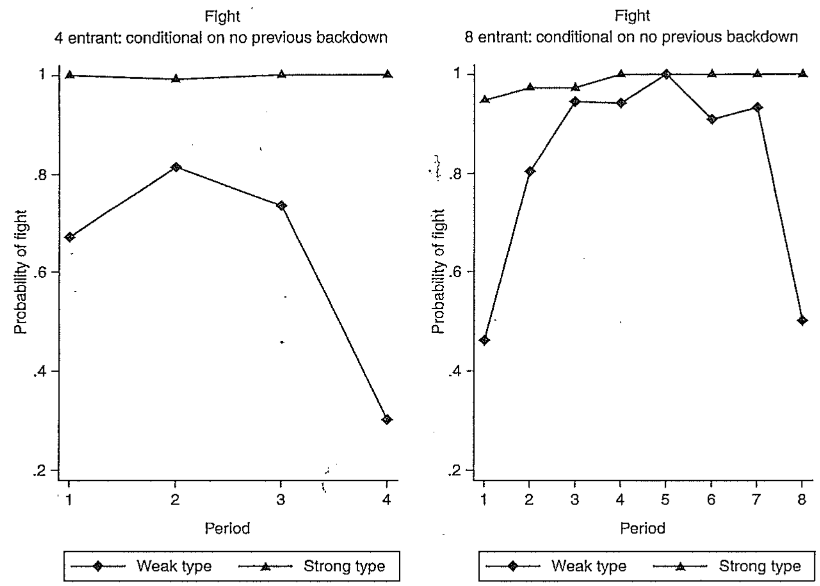

FIGURE 3.5 Probability of fight: four- and eight-entrant designs

they were strong or weak types. Strong types were expected never to back down; weak types were expected to fight a mixed strategy.

Figure 3.5 reveals that strong defenders behaved exactly as the theory predicted. Strong defenders backed down only once out of 492 times in the four-entrant design, and only 8 out of 352 times in the eight-entrant design. This implies that strong types understood the incentive structure for being tough and were behaving exactly according to type.

Weak defenders, however, did not act exactly according to equilibrium predictions. In the eight-entrant design, weak defenders should have fought in periods one, two, and three, and should only have begun to play a mixed strategy in period four. Yet Figure 3.5 shows that only about 45 percent of weak defenders chose to fight in the very first period. Weak defenders acquiesced to very early movers at a far higher rate than they should have to maximize payoffs.

Weak defenders also behaved somewhat differently than expected in the four-entrant design. Here, weak defenders were

expected to play a mixed strategy from the very first period, with fight rates declining from .52, .30, .13, to 0 in the last round of play. Subjects, however, did not decrease their willingness to fight until after the second period, with a sizable decline from the third to fourth period. Hence we find mixed support for hypothesis $H2_D$. Weak defenders did not always fight prior to reaching the mixed strategy stage (in the eight-entrant case). The frequency of fighting, however, declined in the last two periods of the four-entrant design and the last period of the eight-entrant design. This suggests that at least at the end weak defenders came to appreciate the declining value of holding on to a reputation for fighting.

Finally, hypothesis $H3_D$ (which predicted that defenders should resist more in the eight-entrant design in periods 1–4) finds mixed support. In period one, defenders are *more* likely to resist in the four-entrant design. In the second period there is no discernible difference, and in the third and fourth periods defenders were significantly more likely to resist. Thus in two out of the four periods we find support for hypothesis $H3_D$, but in one period we find evidence that rejects the hypothesis.

Why did so many weak defenders give in to the first entrant when they knew they were facing eight entrants in total? And why did so many weak defenders choose to fight against the first three entrants even though they knew they would face only four entrants? It could be that defenders care less about the total number of entrants they are facing than about the number of remaining entrants. In both the four- and eight-player games, the most extreme change in defender behavior occurred at exactly the same time – in the final play of the game. This suggests that some weak defenders may be extremely averse to losing their reputation, only backing down when they know this behavior will not be observed by *any* future players. It also suggests that governments are likely to be more interested in the number of remaining ethnic groups than in the absolute number that exist.

This, however, does not explain why so many weak defenders facing eight entrants would choose to accommodate the first challenger. This is the more puzzling behavior revealed by the experiment. One explanation is that defenders, like entrants, require time to figure out how the game is played. If this is true, then these high rates of accommodation should decline as learning takes place. A second explanation is that these findings are caused by a sub-set of weak defenders who choose to accommodate every entrant no matter what. It is possible, for example, that some defenders simply will not engage in deception, or refuse to manipulate fellow participants. If this is the case, then certain individuals are driving the unexpectedly high rate of early accommodation. Both explanations will be explored more carefully below.

LEARNING

Figures 3.6 and 3.7 investigate learning effects that may be taking place in the course of the experiment. Is the high rate of entry by early challengers, and the relatively low rate of fighting by defenders, the result of inexperience? In order to determine this, we disaggregate the data, showing the choices subjects made each time they repeated the experiment. This allows us to investigate whether the subject's increasing familiarity with the strategic context drives behavior toward or away from equilibrium predictions. It also lets us evaluate whether our decision to aggregate these "repetitions" and report them together was an appropriate one.[22]

Figure 3.6 reveals that entrants did not change their behavior significantly across repetitions of the experiment. In both the four- and the eight-entrant design, the general trend was several periods of declining entry rates against a defender who had not yet backed down, leading to relatively stable entry probabilities of around .45 and .2

[22] The cost is a smaller sample size that reduces our ability to make tight point predictions from the data. We do not model any sort of learning dynamics in our theoretical model, though it is common practice in the experimental literature to compare the behavior of experienced and inexperienced subjects. Indeed, this reflects the notion that a game-theoretic 'equilibrium' is the result of a learning process (Fudenberg and Levine 1998).

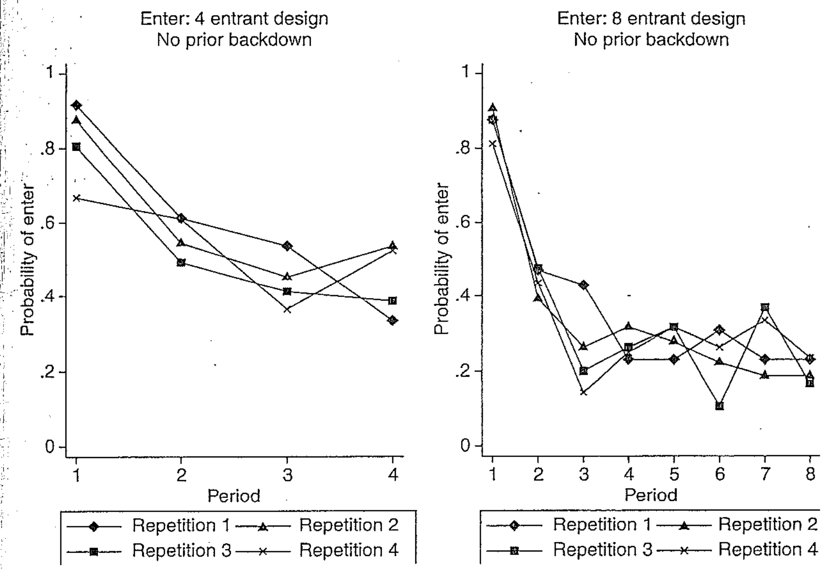

FIGURE 3.6 Probability of enter: four- and eight-entrant designs

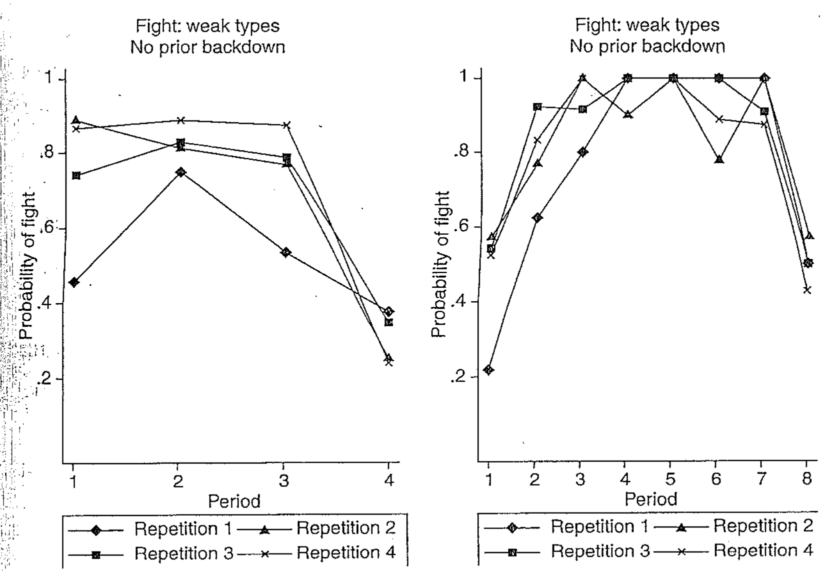

FIGURE 3.7 Probability of fight: four- and eight-entrant designs

respectively. The only significant change is that entrants in the four-person design were notably less likely to enter when they had the greatest amount of experience. This suggests that subjects had learned that defenders in the four-person design frequently chose to fight in the first round but that significant changes in behavior did not occur.

Defenders, however, did alter their behavior as they gained more experience with the game. Figure 3.7 shows how weak defenders behaved across four runs of the experiment. In both the four- and eight-entrant design, weak defenders invested more heavily in reputation over time. This increase took place primarily in the first three periods, with most change taking place between the first and subsequent repetitions of the experiment. Interestingly, this increase was even more pronounced in the four-entrant design. Thus, defenders in both short and long games were learning that fighting early challengers paid off.

Still, it is puzzling why some defenders insisted on backing down to the first challenger even after gaining experience with the game.[23] In an attempt to explain this, we examined two types of individual level data. Since defenders repeated the same sequence of decisions four or eight times, we looked at the percentage of times each subject chose to fight in more than half the periods.[24] Figure 3.8 shows that many weak subjects fell into one of two groups: they either regularly backed down or they regularly fought, with fewer playing a more mixed strategy.

Subjects who were willing to back down in the first period were much more likely to continue to back down in the future. This confirms that a certain type of individual – one who always acquiesces – is at least partly responsible for the unexpected results, making it

[23] The previous analysis pools together all subjects and does not consider any individual differences.

[24] This decision to fight, unlike in previous analyses, was unconditional on what the entrant chose and simply represents their decision to fight following the statement "if the first mover chooses (to enter) what will you do?"

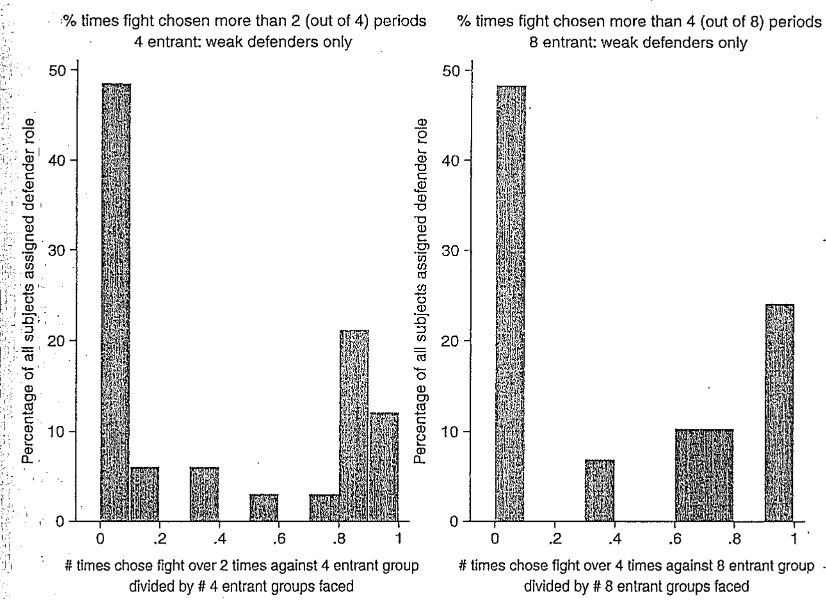

FIGURE 3.8 Percentage of times entrants chose to fight: four- and eight-entrant designs

appear as if all weak defenders are accommodating at a high rate in the very first round when it is actually a sub-set of individuals who are doing so. In the absence of these individuals, the overall rate of fighting in early rounds would be much higher.

Exit interviews also provided insights into why some defenders chose to play consistently sub-optimally. Responses suggest that a certain sub-set of defenders did not appear to understand the strategic logic underlying the reputation building game.[25] The following two responses are indicative of this kind of mistake:

> If you are a weak type, always acquiesce. If you are strong, always fight. It guarantees the most points regardless of what the first mover does. Play the game the same way throughout.

[25] This misunderstanding is not entirely surprising since our instructions were designed not to reveal this logic in order to produce a harder test of the theory.

> Choose the option that gives you more points. If you are a weak type, acquiesce every move. Even if the first mover chooses not to challenge you can't lose. It will get you more money.

But what about the four-person design? Surprisingly, subjects in the four-person design also invested heavily in reputation building in the second and third round even though they would have benefited by being more conciliatory. Weak defenders in these cases did not appear to learn that the costs of fighting were not offset by future gains for having a tough reputation. Why did subjects in the four-person design move further away from equilibrium predictions over time?

One answer is that weak defenders are more concerned with the number of remaining entrants than with the absolute number of entrants. A closer look at Figure 3.6 reveals that defenders in both the four- and eight-person experiments were fighting at about the same rate against the last four entrants. Thus, weak defenders were learning that there were reputational gains to be made from fighting early entrants even in shorter games. As one subject wrote:

> Fight at the beginning of each four period round to fool the other player into thinking you're tough so they will choose not to challenge, and then revert to acquiescing the following three periods. This can work out better than consistently acquiescing depending on how the first movers react. I didn't think of the strategy of faking out the first mover until about halfway through the experiment.

Weak defenders clearly learned that even entrants in the shorter four-person games could be deterred.

The fact that subjects in the four-person designs behaved increasingly tough over time reveals that reputation building dynamics will emerge even in relatively short games. Four iterations were clearly sufficient to generate incentives to be tough. It also suggests that reputation building might still emerge in situations where only

two iterations are expected; weak defenders were willing to continue to invest in reputation building through the second to last play of the game. Additional research will be required to tease out what the critical number of iterations may be for reputation building to become rational in the laboratory.[26]

DISCUSSION AND CONCLUSION

This chapter investigated two fundamental predictions of the chain store model. The first had to do with inferences that were likely to be drawn if a defender backed down against a challenger. The second had to do with the number of challengers and the influence this would have on both defender and entrant behavior.

The experiments reveal that individuals will generally react to incentives to build reputation in the ways the theory predicts. The decision by a defender to grant concessions did appear to send a strong signal to entrants that the defender was weak and did result in a significantly higher number of challenges. Entrants in both the four-person and eight-person designs challenged the defender at a dramatically higher rate if they observed the defender backing down. The defender also seemed particularly worried about sending this signal, often refusing to accommodate until the very last round. Thus, strong inferences were being taken from the past behavior of the defender.

Defender and entrant behavior, however, deviated from equilibrium predictions in two important ways. The most prominent pattern had to do with the willingness of entrants to challenge in early periods, and the willingness of defenders to back down during this time. This was true in both the four- and eight-person games. Although the experiment offers no definitive evidence why individuals behaved this way, individual level data suggest that this result

[26] As mentioned in Chapter 2, we believe that this number is likely to vary not only with the number of future challenges, but with the value of their demands and the strength of their forces.

is driven in part by a sub-set of the defenders who chose to remain true to type despite monetary incentives to the contrary. Entrants then exploited these defenders by "attacking" at an unexpectedly high rate early on.

This seemingly "irrational" play by some of our subjects suggests new and productive avenues for future research. Theoretically, much can be learned from the field of behavioral economics where new equilibrium concepts such as quantal response equilibrium (QRE) and cognitive hierarchy equilibrium (CHE) are increasingly being used (Camerer *et al.* 2004, McKelvey and Palfrey 1998). QRE, for example, may help to explain one of the more interesting findings to come from our experiments – why early entrants would challenge at such a high rate. According to QRE, subjects understand that not all players behave perfectly rationally all the time and they take this "irrationality" into account when determining their strategy. This appears to be what we observed with early entrants. Early entrants challenged at a high rate, but this was quite rational given how often defenders were willing to acquiesce to these early movers.[27] Thus, the "best-response" a player can make will not be the unique sequential equilibrium discussed above, but a response that understands that not all players will behave perfectly rationally all the time.[28]

These results suggest two potential modifications to our expectations about how government and ethnic group leaders are likely to behave when engaged in self-determination disputes. First,

[27] The high rates of entry in early periods might also reflect social comparison issues. If an entrant does not challenge, they get 95 points whereas the defender gets 300 points. If the entrant enters, the point 'spread' is much smaller whether or not the defender fights, or whether it is committed. Thus entry may serve to equalize earnings for both groups. There was some evidence of this in our post-experiment polls, as some entrants expressed frustration that the experiment favored defenders.

[28] Unfortunately, analytical QRE solutions have only been worked out in much simpler games that do not have the complex repetition and updating dynamic in our model. Computer-assisted numerical approaches are not able to account for the updating central to our interests (e.g., using the software package 'Gambit'). This remains an active area of inquiry.

they suggest that most government leaders will care deeply about building a reputation for toughness, and that they are likely to care even when relatively few ethnic groups exist. The findings also suggest that the number of remaining challengers should be more important to governments than the absolute number of potential challengers, and that most governments should choose to fight all but the last remaining ethnic groups.

Finally, the results also suggest that *some* government leaders will be unexpectedly conciliatory and will accommodate any ethnic group that challenges. Knowing that a certain percent of defenders will always concede should, in turn, encourage more ethnic groups to challenge early in a leader's tenure. This may help to explain why some ethnic groups, such as the Slovenes in the former Yugoslavia, are willing to challenge so early.

In the end, the laboratory experiments confirm important parts of the reputation theory, but also reveal where human beings are likely to deviate from existing expectations about rational behavior. Most individuals do understand the logic behind reputation building and use this strategy to their advantage. But the experiments reveal that some people do not, and it is the behavior of these individuals that behavioral economics is trying to explain. In the meantime, we use the insights garnered from these experiments to modify our expectations of what we should observe in the real world, and use Chapters 4 and 5 to test whether the theory holds up in the more complex world of government and minority relations.