Time and Punishment: Blame and Concession in Political Standoffs*

Ian Palmer Cook†
Jonathan Woon‡

March 28, 2016

Abstract

Political negotiation frequently looks like two sides staring each other down, waiting for the other to blink. In these showdowns, neither side wishes to concede, claiming that doing so would incur the wrath of voters. Whether this consideration of potential punishment influences behavior during stalemates is not well understood, and little theory or evidence exists to explain how voters allocate blame for different outcomes. We conduct a laboratory experiment to investigate two interrelated questions: how does anticipation of blame drive behavior, and how do observers with a stake in the outcome allocate blame? In our experiment, we adopt a dynamic war of attrition to model a negotiating situation in which concession time is the key choice variable, and our design compares versions of the game with and without an observer (whose payoffs depend on the outcome and who can punish the players). We find that the presence of an observer has little effect on standoff outcomes but appears to shorten the duration of standoffs. We also find that observers tend to punish the winning player, which is qualitatively consistent with a rational or instrumental form of punishment, but the amount of punishment is less than optimal. Negotiators, like politicians, may not adequately anticipate when, or how much, they will be blamed, preventing them from altering their choices and achieving better outcomes.

*Thanks to Catherine Hafer, Kristin Kanthak, George Krause, and Victoria Shineman for helpful comments and discussion. A prior version was presented at the 7^{th} Annual NYU-CESS Experimental Political Science Conference.

†Ph.D. Candidate, Department of Political Science, University of Pittsburgh. Corresponding author: ipc4@pitt.edu.

‡Associate Professor, Department of Political Science, and Faculty, Pittsburgh Experimental Economics Laboratory, University of Pittsburgh

> "Our message to the United States Senate is real simple: The American people don't want the government shut down, and they don't want Obamacare."
>
> ---
>
> Rep. John Boehner

> "I want to be absolutely crystal clear: Any bill that defunds Obamacare is dead. Dead."
>
> ---
>
> Sen. Harry Reid

Political negotiation often resembles a staring contest, rather than the back–and–forth offer–counteroffer process usually associated with deal-making and compromise.[1] Consider recent struggles over U.S. fiscal policy. In September 2013, congressional Republicans demanded the elimination of funding for the Affordable Care Act as a condition of passing continuing appropriations to keep the government operating. Democrats refused to back down, and a 17 day government shutdown ensued. During the shutdown, polls indicated that the public largely blamed Republicans and eventually the Republican leadership capitulated.[2] Four months later, having reasoned that allowing the shutdown had been a mistake, Republicans would not be so intransigent.[3] Facing a looming deadline over raising the debt limit, they acceded to Democratic demands well ahead of the potential default, and the public hardly noticed.[4] Similar events unfolded in 1995 between Newt Gingrich and Bill Clinton. These events suggest that incurring blame is costly for politicians, and that the anticipation or avoidance of blame can feature prominently in strategic calculations.

[1] We generally avoid the term "bargaining" throughout the paper. While standoffs occur during a bargaining process, our interest is on the impasse and potential failure of the bargaining event, rather than choices of offers during bargaining. Both our model and experiment are sufficiently different from the usual models of bargaining that we want to keep the distinction clear for the reader. We make an exception in the discussion of prior literature, where commonalities between our study and work on bargaining is clear.

[2] Andrew Dugan, "Republican Party Favorability Sinks to Record Low." *Gallup Politics*, 9 October, 2013. Scott Clement, "Republicans are losing the shutdown blame game", *The Fix*, 4 October, 2013.

[3] John Breshnahan, Manu Raju, Jake Sherman, and Carie Budoff Brown, "Anatomy of a shutdown", *Politico*, 18 October, 2013.

[4] Paul Kane, Robert Costa, Ed O'Keefe, "House passes 'clean' debt-ceiling bill, ending two-week showdown", *Washington Post*, 11 February, 2014.

How do observers of political stalemate allocate blame for outcomes, and how does the expectation of blame affect the behavior of negotiators? Answering these questions is important for understanding how citizens can exert influence over their representatives through non-electoral means. To study these questions we conduct an experiment that contrasts standoffs with and without an audience. In our setup, subjects play a war of attrition game in which the only choice is when to back down and waiting is costly. A single-shot game is used to isolate the incentives for allocating blame for the outcome from concerns about discounting, reputation effects, and potential unravelling results. When we add an audience, we operationalize blame by allowing the observer to punish the standoff contestants. The war of attrition framework captures the essential features of standoffs outlined in the example above. It is also important for our purposes because ex ante negotiating power is symmetric, permitting us to investigate how observers ascribe blame as a function purely of outcomes, free of considerations of structural power.

This paper proceeds as follows. First, we emphasize the importance of studying blame in negotiations and standoffs, then demonstrate that our approach is both novel and substantively appropriate. Next, we describe the design and procedures of the experiment. From this we describe what payoff maximizing behavior predicts for Staring Contest outcomes and the use of blame. Briefly, if observers engage in rational punishment and contestants correctly anticipate this, then the presence of an observer in the Staring Contest will shorten the time players hold out before conceding. In addition, the contestant whose preferences are aligned with the observer should win more often. This occurs because the observer adopts a strategy that punishes the winning player, thereby reducing delay. The empirical results provide mixed support for the predictions of rational play; behavioral models that incorporate preferences for equity of winnings may explain some of the gap. The observer frequently uses blame to target the winning player, and contestants decrease their waiting times in the presence of an observer. Outcomes, however, change only slightly, if at all. Contestants, it seems, inadequately anticipate how they will be punished and fail to take action to improve their payoffs. We conclude with a discussion of these findings and their general implications.

1 The Politics of Blame

We consider blame to be a form of punishment citizens use to register their displeasure concerning some event or action. Political decisionmakers seek to avoid punishment at all levels, from elites (Weaver, 1986; Weale, 2002; Hood, 2010) to street-level bureaucrats (Lipsky, 1980; Brehm and Gates, 1997). Doling out punishment to sitting politicians entices some voters to the ballot box (Peffley, 1984; Iyengar, 1989; Brown, 2010). Negative assessments of political culpability, such as opinion polls, are weak forms of blame, but arise in numerous contexts (e.g., Bengtsson, 2004; Malhotra and Kuo, 2007; Alcañiz and Hellwig, 2011; De Vries and Hobolt, 2012). We build on prior research about responsibility attribution, as well as work on strategic interactions between bargainers and an audience.

Attributing blame requires an observer to make a valid link between the causal drivers and the outcome. Citizens have weak, and malleable, understanding of these links when considering large scale failures (Malhotra and Kuo, 2007). Indeed, they mostly rely on partisan identification when deciding who is most responsible for economic conditions or policies (Brown, 2010). Blame is, however, more nuanced when citizens have better information (De Vries and Hobolt, 2012). In contrast to our setting, these studies present large, often highly complicated problems (massive environmental disasters and macroeconomic conditions) to voters with varying internal motivations for devoting attention and thought. We isolate these competing pressures in the experiment by providing complete information and clearly delineating incentives, and making the causal link transparent.

Blame, as a logical subset of responsibility (Lagnado and Channon, 2008; Shaver, 1985), carries similar importance to the study of economic voting. However, while theories of prospective and retrospective voting entail the attribution of responsibility for economic conditions (Lewis-Beck and Paldam, 2000; Lewis-Beck and Stegmaier, 2000; Paldam, 2008; Hellwig, 2010), they do not address how voters *apportion blame* for producing those conditions. Psychologists define blame as the impulse to punish an actor for a negative outcome over which the actor has some direct control (Brewin and Shapiro, 1984; Weiner, 1995; Alicke, 2000).[5] The distinction between responsibility without blame and responsibility with blame is important because it implies potentially different voter behavior. For example, voters may hold the president responsible for employment levels, but do not *blame* him for conditions inherited from the last administration. Voters may, however, blame a political party that blocked job-creating legislation, then punish them at election time. Treating blame as a distinct concept provides a richer view of voter behavior.

Blame and reputation are interrelated, but not synonymous; distinct models of each are necessary. Repeated games often model reputational concerns as beliefs about an opponent's promised future actions, given the opponent's history of behavior, or unobserverable payoff-relevant types (Kreps and Wilson, 1982; Kreps et al., 1982; Roth, 1985; Celetani et al., 1996; Abreu and Gul, 2000; Embrey, Fréchette and Lehrer, 2014). These models do not address punishment as a distinct mechanism, and reputational concerns arise only between bargainers, since no audience is present. Another form of reputation arises in games of incomplete information, such as in Groseclose and McCarty (2001), who incorporate both blame and an audience in a model of veto bargaining. Voters, as the audience, apply blame by lowering approval ratings for the president if he vetoes a certain bill. In this model, the voters assign both full responsibility, and thus full blame, to only one side of what is a strategic interaction. In contrast, our experiment allows the audience to blame neither, one, or both parties according to their assessment of both actors' choices. International

[5] This is a counterfactual-based definition of causality. A caused B (in whole or part) if B would not have happened without A's presence. The *desire* to punish this causal connection is sufficient for psychology, though we choose to make it tangible.

relations studies of bargaining often include audiences that can impose a cost on political leaders as a function of the bargaining outcome (Fearon, 1994; Tomz, 2007*b*,*a*; Weeks, 2008). Moreover, we follow Fearon (1994) in employing a war of attrition model. Domestic politics, too, are often described as wars of attrition (Wawro and Schickler, 2006), and have been explicitly modeled this way (Kousser and Phillips, 2009; Kovenock and Roberson, 2009). We complement this by making the cost mechanism – a subject of some debate (Schultz, 2001; Trachtenberg, 2012) – explicit.

Our operationalization of blame is similar to the use of punishment in public goods experiments, making them appropriate bases on which to build our examination of standoffs and players' outcome preferences.[6] We choose to make blame costless to the observer, to render the application of punishment independent of the game's structure. That said, experiments report significant use of punishment even when the punisher incurs a cost to do so (Fehr and Schmidt, 2000; Fehr and Gächter, 2002; Masclet et al., 2003). Punishment has proven useful in studying a wide array of social situations, despite debates about what the findings mean (Guala, 2012). Blame may prove a similarly useful experimental tool for understanding political phenomena.

2 Experimental Design and Procedures

To investigate how the presence of an audience and the potential for punishment affects behavior during a standoff, and how an audience might actually punish the standoff contestants, we designed a simple experiment that contrasts standoffs with and without an audience. For our framework, we chose to use a war of attrition game in which waiting is costly. When we add an observer, the observer's payoffs depend on the outcome of the attrition game, and we operationalize "blame" by allowing the observer to punish the players by deducting points.

[6] The general public goods literature is expansive. For a useful summary of punishment and cooperation in experimental settings, see Chaudhuri (2011).

We also vary the alignment of interest between the observer and the players by modeling outcomes as spatial policies and varying the ideal point of the observer.

We chose the war of attrition framework rather than a sequential bargaining game for two reasons. Substantively, we are interested in the duration of standoffs, as we believe that recent episodes of presidential-congressional and inter-cameral negotiating over the federal budget, debt limit, and sequestration seemed to resemble a process in which each side stakes out a position and waits for the other to concede rather than a process of negotiation that leads to compromise outcomes. From a design standpoint, the war of attrition framework allows us to keep proposal power symmetric between the standoff contestants. This is important because we are interested in understanding how observers endogenously assign blame as a function of the choice to prolong or end the stalemate, free of considerations arising from exogenous structural features of a standard bargaining situation (such as proposal power or a first mover advantage).

The experiment utilized a within-subject design and consisted of two parts. In Part 1, subjects played a war of attrition game without an audience, while in Part 2, they played the same game with an observer who was permitted to administer ex post punishment. This ordering follows our theoretical arugment: the first step establishes a baseline against which the behavior in the second part can be compared. Mirroring the theoretical and experimental steps clarifies the latter empirical evaluation, and keeps the comparison of effects clean. We also implemented standard design features to minimize the interaction between rounds, ensuring as much as possible the one-shot nature of the incentives of each play of the game. First, we used anonymous, random rematching of groups between rounds so that subjects never knew which of the other subjects they were paired with. This feature reduces any reputational incentives subjects may have. Second, we randomly selected one round from the entire session to count as payment. This feature eliminates wealth effects and ensures that a subject's payoffs in later rounds are independent of their payoffs in earlier rounds.

In Part 1, subjects played 12 rounds of a two-player war of attrition game without an audience. In our experimental instructions, we referred to the war of attrition as a "Staring Contest Game" to emphasize its dynamic nature and to convey the intuition that the player who waits the longest in the game obtains a better outcome. We can think of the player who "stares" the longest as "winning" the game and the player who "blinks" first as "losing." Although we referred to the game as a Staring Contest, we did not otherwise use the terms "staring," "blinking," "winning," or "losing" to describe the game.

The Staring Contest Game is played by two players, which we designate Player A and Player B. Each subject played 6 rounds as Player A and 6 rounds as Player B. The game lasts for 30 seconds, and each player's only decision is when to "concede" by ending the game. In our graphic interface for the game — programmed in z-tree (Fischbacher, 2007) — each player sees a timer bar that decreases in size at 1-second intervals, and chooses to concede by moving the cursor over the timer bar. We designed the interface in this way so to avoid audible clicks of the mouse that would signal to other subjects when players in other groups conceded. If both players concede at the same time, we use a random tie-breaking rule.

When the Staring Contest is played dynamically, as in our design, we only observe the actual concession time of the game for the losing player. We do not observe the time at which the winning player would have conceded, so in a way, our data are censored. The standard solution to this kind of censoring problem in experiments with sequential games is to use the "strategy method," in which subjects do not play the game dynamically but instead indicate their complete strategies before other players' choices are revealed. But the strategy method removes the dynamic element of the game that interests us. Thus, we opted to use a mixture of dynamic play and the strategy method. At the beginning of each round of the Staring Contest, we asked each player to state an "Intended Stopping Time" (between

0 and 30 seconds), and there was a 1 in 10 chance that we would implement the Intended Stopping Time as the player's actual stopping time. The realization of the Intended Stopping Time was independent across players and periods. We intended the small chance that the Intended Stopping Time would be implemented as a way to make the choice meaningful (rather than completely hypothetical) while ensuring that subjects would play the dynamic version of the game most of the time.

We described the outcomes and payoffs of the game in terms of a one-dimensional spatial model of policy.[7] If Player A wins, the "outcome number" of the game is $a = 10$, while if Player B wins, the outcome is $b = 90$. If neither player concedes, we consider the outcome of the game to have broken down in so far as neither side gets the outcome they prefer despite both holding out.[8] To construct the players' payoffs in the Staring Contest, each player is described as having a "target number," which we can think of as the player's ideal point. We assign A the target number 0, and B the target number 100. If we denote the outcome of the game by $x \in \{a, b, \phi\}$ (where ϕ denotes disagreement) and the game ends at time t (in seconds), then player i's payoff (denominated in points) is given by the function

$$u_i(x,t) = \begin{cases} 350 - |x - \theta_i| - t & \text{if } x \neq \phi \\ 190 & \text{if } x = \phi \end{cases}$$

where θ_i denotes the player's target number. Note that "winning" the Staring Contest results in payoffs between 310 and 340 points, while "losing" results in payoffs between 230 and 260 points. Disagreement is the worst possible outcome, and it is equally "bad" for both players in order to preserve the symmetry of the players' incentives in the game. Substantively, neither side of a political standoff needs to know the actual value of getting one policy or

[7]When we introduce an observer in Part 2, this allows us to manipulate the alignment of preferences between the observer and the two contestants.

[8]We do not use the terms "breakdown", "bargaining failure", "standoff", or any similar term in describing this outcome to subjects. We only refer to a default payoff if neither player concedes.

the other. Instead, these actors only need an ordinal ranking: the party getting exactly the policy they want is the best outcome; any policy between their preference and the other party's is worse; a policy exactly at the other party's preference is worse still; finally, the outcome corresponding to a government shutdown in which *no* policy is enacted, not even the status quo, is the worst possible outcome for either party.

In each round of Part 2, randomly matched groups of three participants played the Staring Contest Game with an audience. We designated two players in each group to be the "contestants" (Players A and B) in the Staring Contest and the third player to be an "observer." Every subject played 6 rounds in Part 2 as Player A, 6 rounds as Player B, and 6 rounds as the observer. Play in each round consisted of two stages. The contestants first played the Staring Contest Game exactly as they did in Part 1. The observers then learned the outcome and their payoffs from the Staring Contest and chose how to allocate "blame" or "punishment" by deducting points from one or both contestants' payoffs. The observer has the ability to allocate punishment to one, both, or neither of the contestants and could deduct any amount of points as long as the sum of deductions was between 0 and 100 points.[9] We chose to make punishment costless rather than costly so we could observe the maximum amount of punishment subjects might be willing to give.

The contestants' payoffs in this version of the game are their Staring Contest payoffs minus the observer's deduction, $u_i(x,t) - d_i$, where d_i is an integer between 0 and 100 and the sum $d_A + d_B$ is between 0 and 100. The observer's payoff is given by the same function $u(x,t)$ that describes the contestant's payoffs in Part 1, except that we varied the ideal point of the observer, $\theta_O \in \{0, 25, 50\}$, in order to vary the alignment of interests between the observer and contestants. When $\theta_O = 50$, the observer's interests are aligned equally with Player A and Player B. When $\theta_O = 25$, the observer's interests are more aligned with Player

[9] Van De Ven and Villeval (2014) develop a deception game with an observer that can reveal a player's lie. Both this model and ours make the player outcomes dependent on the choice of a second-mover audience, though the negotiating parties in our model cannot take unobserved actions.

A than Player B. And when $\theta_O = 0$, then the observer is completely aligned with Player A.[10] This alignment models partisan bias in the audience, doing so only for one side to keep the experimental conditions tractable. Information about the observer's ideal point was announced to all players at the beginning of every round and therefore common knowledge. Each subject played two rounds in each role with each of the possible observer ideal points.

A total of 72 subjects participated in five sessions of the experiment, which took place at ****** .[11] Subjects were recruited from ****** general subject pool, gave informed consent according to standard procedures, and were privately paid in cash at the conclusion of the experiment (including a $5 show-up fee).

3 Theoretical Expectations

In our theoretical analysis, we first consider how payoff-maximizing subjects would play the baseline Staring Contest Game and then consider how the observer's actions might affect concessions and timing. Even though subjects play the game in real time, we can analyze the game as a simultaneous move game in which a player's strategy is the amount of time (in integers) he or she waits until conceding, denoted by t_i for $i \in \{A, B\}$.

In the game without the observer, each player always has an incentive to wait longer than the other. For any strategy for Player B, $t_B < 30$, Player A's best response is to wait a little longer and to concede at any time t_A such that $t_A > t_B$. To see why, note that A's payoff for waiting longer than B is $340 - t_B$ while conceding at $t_A < t_B$ yields a payoff of $260 - t_A$. Since $t_B - t_A$ is at most 30 points and the difference between winning and losing is 80 points, it is better to wait and win the contest than it is to end the game quickly on the losing side

[10] Note that the observer can make conceding the worst possible outcome by deducting 80 points or more from the conceding player.

[11] One subject voluntarily left the experiment during the last rounds of Part 1, which also forced us to involuntarily dismiss additional subjects. We use data from this session in the following results. The number of subjects is reduced to 66 in cases where the excluded subjects did not participate. The basic results do not change if these data are excluded.

(because $340 - t_B > 260 - t_A$). If both players wait until the last second, $t_i = 30$, then the random tie-breaking rule implies an equal likelihood of obtaining a payoff of 310 (from winning) and a payoff of 230 (from losing). The expected payoff when both players wait until the very last second to concede is therefore 270, which is still preferable to conceding immediately and obtaining 260 points. Thus, the Nash equilibrium is for both players to wait until the last second, $t_i = 30$, to concede. Our first prediction is a simple application of this result, provided subjects are risk neutral. However, if subjects are sufficiently risk averse, then they will prefer conceding immediately and receiving a payoff of 260 points with certainty to waiting until the end of the game and receiving a risky outcome with an expected value of 270 points. In this case, we would expect immediate concessions. It then follows that if there is heterogeneity in subjects' risk preferences, we would expect to see a mixture of concession times at the extremes of 0 and 30.

Modeling the Staring Contest as a game of incomplete information would not substantively change these hypotheses. Consider two players with unknown types, defined as their level of risk acceptance. If player A believes her opponent to be more risk-seeking, and thus that Player B would be willing to out-wait her, then it is strictly better for A to concede as soon as possible since waiting is costly and waiting until $t = 30$ results in a toss-up. Conversely, if Player A believes Player B to be less risk-seeking, then at every time t Player A's incentive is to continue waiting until $t = t + 1$ because she expects Player B will concede at t, continuing until $t = 30$. (The argument for Player B's choices are symmetric.) Thus, as with the above model, we expect concessions to occur at the beginning or end of the Staring Contest. We can summarize this expectation as the first hypothesis.

Hypothesis 1 *In the Staring Contest Game without an observer, risk neutral players will wait until the last moment before conceding while sufficiently risk averse players will concede immediately. A mixture of risk neutral and risk averse players will produce waiting times at the extremes of 0 and 30 seconds.*

Now consider the game when an observer is present. In the spirit of backward induction, we formulate expectations about what a payoff-maximizing observer would do and then discuss how contestants would change their behavior in anticipation of the observer's response. Because the observer moves last and takes an action that has no effect on her own payoff, she will be indifferent between allocations of punishment along the path of play. In other words, any allocation of punishment can be supported in equilibrium. However, the observer is not indifferent between all possible outcomes and can design a punishment strategy that, if rationally anticipated by the contestants, induces the best possible Staring Contest outcome for herself. We specifically do not adopt any refinement of the possible equilibra, but instead choose to concentrate on the equilibrium supported by the lowest amount of punishment that would induce a change in behavior for the contestants (again, assuming a strict notion of rationality). Any greater amount of punishment than what we choose would induce the same behavior on the part of the contestants, and the observed result would be identical for any level of punishment above the minimum we choose. In effect, we face a problem of observational equivalence, and choose to address it by setting the punishment high enough to induce a change in contestant choice of stopping time, but no higher.

The level of blame used by an observer is endogenous and accounts for expressions of partisan bias. The extent to which voters strictly side with one party or another, however, is unclear since political ideology and party identification are not synonymous for the modern U.S. electorate (Treier and Hillygus, 2009). The gap between ideology and identification is naturally embodied in our experiment: Observers can allocate blame based on ideological distance between the outcome and their preference, party loyalty to Player A regardless of the strength of affiliation, or some combination of both.

When the observer's ideal point is $\theta_O = 50$, she is indifferent between the outcomes $a = 10$ and $b = 90$ and thus cares only about minimizing the duration of the Staring Contest. That is, she does not care which contestant wins as long as $t_A = 0$ or $t_B = 0$. The observer can force the contestants to end the game immediately by deducting at least 80 points from the winning player. To see how this encourages the desired behavior, note that deducting 80 points from the winner implies that the winner's payoff will be $260 - t$. Note also that the loser's payoff is $260-t$, so the observer's strategy effectively makes the contestants indifferent between winning and losing.[12] If contestants rationally anticipate the observer's strategy, their best response is to end the game at $t = 0$, yielding the maximum payoff of 260 points.

In rounds where the observer's ideal point is either $\theta_O = 25$ or $\theta_O = 0$, her ideal outcome is for Player B to concede at $t = 0$ and for Player A to win the contest, which yields the outcome $a = 10$. The observer can therefore design a strategy that encourages Player B to concede at $t_B = 0$ by deducting at least 80 points from B if and only if B wins the Staring Contest. To illustrate how this strategy works, consider any outcome where A is the first to concede at $t_A > 0$. If the observer deducts the maximum 100 points from Player B, then B's payoff would be $240 - t_A$. By deviating to some other time $t_B < t_A$, Player B increases his payoff to $260 - t_B$ and can maximize his payoff by stopping immediately at $t_B = 0$. To see that the stopping times $t_B = 0$ and $t_A > 0$ constitute an equilibrium given the observer's punishment strategy, note that this yields payoffs of 340 for Player A and 260 for Player B. Player A will not deviate to $t_A = 0$ and risk obtaining the lower payoff. Similarly, Player B will not deviate to any $t_B > t_A$ because doing so would incur the observer's punishment and yield a lower payoff of $240 - t_B$.

The following hypotheses summarize the effects of introducing an audience on contestants' behavior and the ways in which we expect payoff-maximizing observers to play. In terms of standoff outcomes, the effect of the audience should be to decrease the observed and intended stopping times (Hypothesis 2). When the observer's preferences align with Player A's, we also expect to see standoff behavior and outcomes in which Player A wins the Staring Contest more often than Player B (Hypothesis 3). In terms of blame and punishment,

[12] For this strategy to work, the observer must also punish the winning player even if $t = 0$ because otherwise each contestant has an incentive to deviate to $t_i > 0$ given that the other contestant $j \neq i$ concedes immediately at $t_j = 0$.

observers will generally punish the winning player in order to create incentives that minimize delay in standoffs (Hypothesis 4). Finally, observers will direct punishment towards Player B when she prefers Player A win the Staring Contest (Hypothesis 5).

Hypothesis 2 *Waiting times will be shorter and the Staring Contest is more likely to end immediately with an observer than without an observer.*

Hypothesis 3 *As the distance between the observer's ideal point and Player A's ideal point decreases, Player A's waiting time will increase, Player B's waiting time will decrease, and Player A is more likely to win the Staring Contest.*

Hypothesis 4 *When $\theta_O = 50$, the observer is more likely to punish the winning player than the losing player.*

Hypothesis 5 *When $\theta_O = 25$ or $\theta_O = 0$, the observer is more likely to punish Player B for winning than Player A.*

4 Findings

The results of the experiment support several of our hypotheses. We find support for the two major aspects of the theoretical predictions regarding the outcomes of the Staring Contest and the amount of punishment levied by the observer. In our analysis, we also classify the types of strategies observers use and assess how behavior correlates with individual characteristics.

4.1 Outcomes

In the baseline Staring Contest without an observer, 94% of plays of the game result in a concession. Overall, the outcomes are close to evenly split between Player A winning (50%) and Player B winning (44%), but with a slight advantage for Player A in terms of outcomes. Figure 1 shows the distribution of actual and intended concession times in the baseline Staring Contest without an observer.[13] Actual stopping times follow a bimodal distribution, with the majority of contests ending around $t = 0$ and $t = 30$. Note that the distributions are bimodal in *all* treatments; we thus do not use the "average" duration of the Staring Contest to compare the results of non-observer and observer conditions. The game ends immediately (at $t = 0$) in 8% of rounds played and at the last possible second (at $t = 30$) in 9% of rounds played. If we allow for some error in waiting times, we find that 25% of rounds played end within the first 5 seconds while 35% end within the last 5 seconds. Intended waiting times are roughly similar. Six percent of subjects indicated they intended to concede immediately (at $t = 0$), while 16% indicated they intended to wait until the very end (at $t = 30$). Further, 61% of all intended stopping times fall within the first five seconds (19%) or the last five seconds (42%).[14] Thus, without an observer, the majority of games ended—or subjects stated their willingness to end the game—at the very begnning or end of the waiting period, generally consistent with the prediction stated in Hypothesis 2.[15]

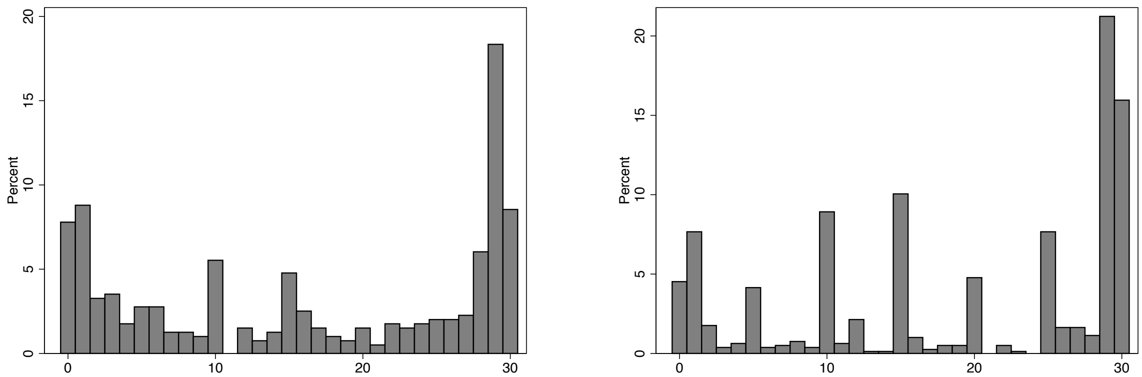
Turning now to the effects of introducing the observer on Staring Contest outcomes, we find mixed support for our hypotheses. Table 1 presents the average actual and intended stopping times by observer condition. Although there is a very slight decrease in actual waiting times, the average waiting times are nearly identical across treatments and between the contestants. These results suggest that introducing an observer has little effect on the duration of the standoff. In every condition, we also note that the intended stopping times are always longer than the actual waiting times, and Player A's intended stopping times are longer than Player B's. For a more rigorous test of whether the observer affects standoff

[13]The Pearson correlation coefficient for actual and intended stopping times is $\rho = 0.38$. This is not an especially high correlation, so we continue to present results for both measures in this section. The results for both are substantively very similar.

[14]We choose this limit to account for human response time during the contest. Subjects may intend to quite immediately, but may err in their ability to do so at $t = 0$. Five seconds accounts for the possibility of physical error (mis-handling the concession function), or even a lack of attention.

[15]Subjects do exhibit some learning effects as they play more rounds. Specifically, they are more likely to concede near $t = 30$ the more rounds they play. However, we believe this has no substantive effect on our findings. Details can be found in Appendix A.

Figure 1: **Actual Stopping Times Bimodal, Intended Times Cluster**



(a) **Actual Stopping Times** (b) **Intended Stopping Times**

durations, we estimate regressions of stopping times on a set of treatment variables with subject fixed effects. The results, presented in Table 2, indeed suggest that the observer has no statistically significant effect on average waiting times. The exception is a statistically significant decrease in Player A's intended waiting time when the observer's ideal point is 0, but this effect is the opposite what the theoretical analysis predicts for Player A. Our analysis of average waiting times lends no support to Hypothesis 2 or Hypothesis 3.

When we look at the distribution of waiting times instead of the average in the observer conditions, shown in Figure 2, there do appear to be differences between the conditions in terms of the proportion of early and late concessions. The number of players conceding immediately (or very early) is higher under all observer treatments than under the no observer treatment. In rounds when there is an observer, the proportion of rounds with immediate concessions ($t = 0$) more than doubles, increasing from 8% to $18-22\%$. There is also an increase in terms of the proportion of rounds that end early (within the first 5 seconds) from 25% in the baseline to $32-37\%$ with an observer.[16] Table 3 shows that these effects are statistically significant. The first column presents the estimates for a probit model in which

[16] Conducting Kolmogorov-Smirnov equality of distribution tests shows that the distribution under the No Observer treatment is statistically significantly different from each of the observer treatments ($p < 0.01$ for each treatment).

Table 1: **Average Actual and Intended Stopping Times by Treatment**

		Actual	**Intended Player A**	**Intended Player B**
No Observer	Mean	16.4	20	18
	N	398	426	426
Observer 50	Mean	15.7	19.7	19
	N	128	132	132
Observer 25	Mean	14.2	19.6	17.5
	N	122	132	132
Observer 0	Mean	14.8	18.7	18.2
	N	125	132	132
Total	Mean	15.7	19.7	18.1
	N	773	822	822

Table 2: **Effects of Treatment on Contestant Choice of Stopping Times**

	(1) **Actual**	(2) **Intended Player A**	(3) **Intended Player B**
Observer 0	-1.62	-1.63*	-0.15
	(1.22)	(0.76)	(0.82)
Observer 25	-2.14	-0.73	-0.86
	(1.24)	(0.76)	(0.82)
Observer 50	-0.73	-0.65	0.60
	(1.21)	(0.76)	(0.82)
Constant	16.39***	20.15***	18.15***
	(0.60)	(0.37)	(0.40)
R^2	0.01	0.01	0.00
Observations	773	822	822

Standard errors in parentheses.
No Observer treatment omitted as reference category.
Models (2) and (3) use subject fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the dependent variable is an indicator for whether the game ends immediately at $t = 0$; the probability the game ends immediately is significantly higher for each of the observer conditions. The dependent variable for the second column is an indicator for whether the game ends early (within the first 5 seconds); while the coefficients are all positive, the effect of the observer when $\theta_O = 25$ is significant at the 0.05 level, the effect is significant at the 0.10 level for $\theta = 0$. The dependent variable for the third column is an indicator for whether the game ended late (the last five seconds); there are no significant difference from the No Observer treatment. The dependent variable for the final model is an indicator for whether the game stopped at the last possible moment ($t = 30$). Note that the coefficients in column four are negative, and the effect of the observer when $\theta_O = 50$ is significant at the 0.05 level, indicating that, under this treatment, the game is less likely to end at the last second. In contrast to the analysis of average waiting times, our analysis of immediate concessions suggests some support for Hypothesis 2.

Table 3: **Effects of Treatment on Likelihood of an Extreme Stopping Time**

	(1) **Immediately**	(2) **First 5 Seconds**	(3) **Last 5 Seconds**	(4) **Last Second**
Observer 50	0.50^{**}	0.20	0.09	-0.33^{*}
	(0.16)	(0.13)	(0.13)	(0.16)
Observer 25	0.41^{*}	0.34^{*}	-0.05	-0.12
	(0.17)	(0.13)	(0.13)	(0.18)
Observer 0	0.66^{***}	0.23^{+}	-0.01	-0.27
	(0.16)	(0.13)	(0.13)	(0.17)
Constant	-1.42^{***}	-0.67^{***}	-0.33^{***}	1.37^{***}
	(0.09)	(0.07)	(0.06)	(0.09)
Observations	773	773	773	773
Pseudo R^2	0.04	0.01	0.00	0.01

Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 2: **Distribution of Actual Stopping Times**

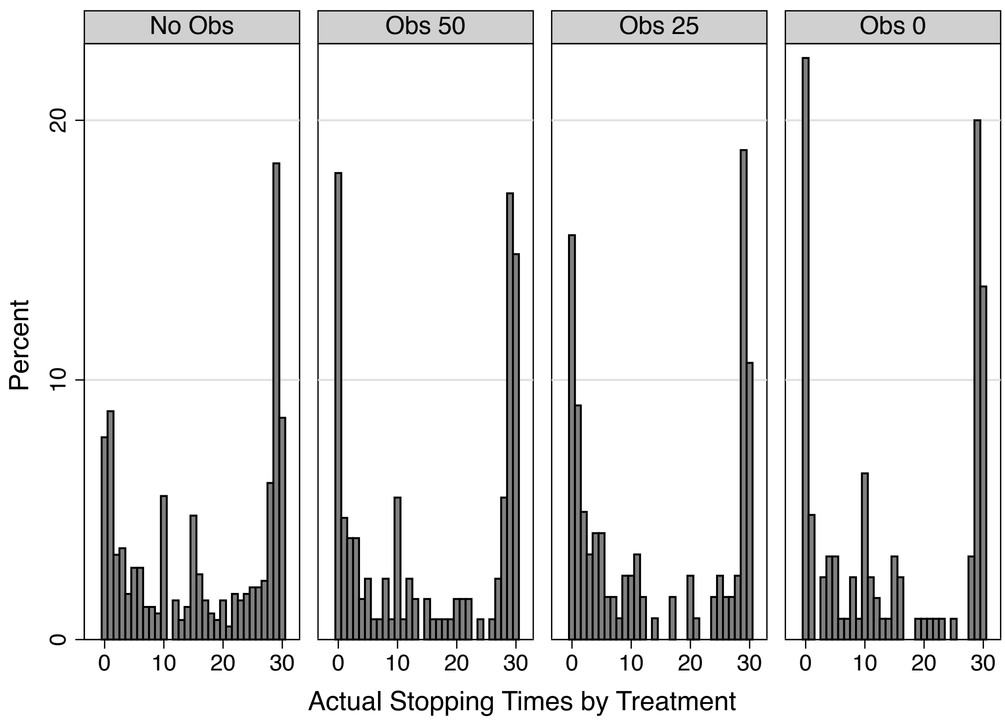
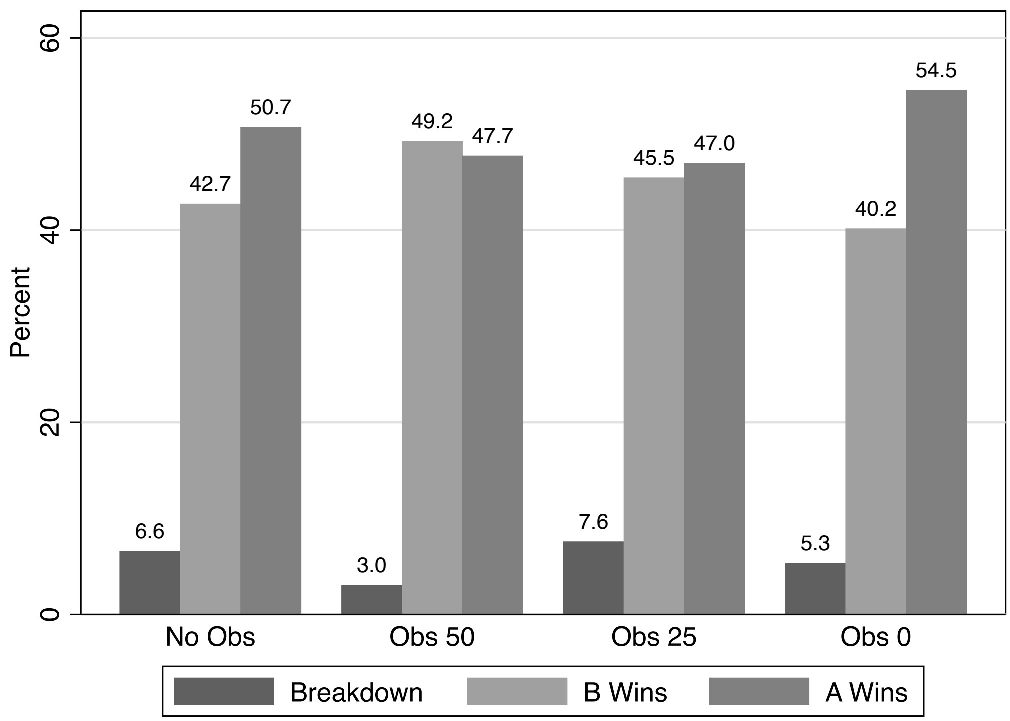


Figure 3 shows the distribution of outcomes from the Staring Contest by observer condition. The most notable result is that the presence of an observer appears to have little, if any, influence on the outcomes of the contest. There appear to be slight differences in the proportion of rounds that end in breakdown, but since there is no clear pattern, these may simply be noise. In terms of whether Player A or Player B wins the contest with an observer, it does appear that Player A is less likely to concede as the alignment between the observer and Player A increases. Furthermore, Player A wins the contest more frequently when the observer's ideal point is $\theta_O = 0$ than in any other conditions. However, these differences are not statistically significant, so there is little support for Hypothesis 3.

Figure 3: **Distribution of Outcomes by Treatment**



To summarize, we find modest effects of the presence of an observer on the duration of the standoff. The basic outcomes of the staring contest (who wins or loses) appear to be insensitive to the addition of an observer. However, the distribution of stopping times changes: when an observer is present, subjects are more likely to end the game immediately.

4.2 Punishment

Broadly speaking, we find that observers allocate punishment in ways consistent with our theoretical expectations. They appear to punish a contestant for winning the contest more than they punish a contestant for losing, and they also appear to punish Player B for winning more often than they punish Player A. We also find evidence that observers punish both contestants heavily for breakdown.[17]

In Table 4 we present the average number of points the observer deducts from each contestant by outcome and treatment condition. The pattern of behavior appears to lend support to both Hypothesis 4 and Hypothesis 5. First, we see that observers punish the winning player more than they punish the losing player. When Player A wins, the observer deducts an average of $4.4 - 7.3$ points from A and $15.4 - 27$ points from B. Similarly, when Player B wins the Staring Contest, the observer deducts an average of $6.4 - 22.5$ points from A and $23 - 40$ points from B. These patterns are consistent with Hypothesis 4. Second, the deductions from Player B appear to be much more sensitive to the outcome than deductions from Player A. For example, when the observer's ideal point is $\theta_O = 0$, the observer deducts nearly 10 times more points from Player B when B wins as when A wins. In contrast, the observer deducts about the same number of points from A for both outcomes. Table 4 also suggests that Player B's punishment for winning increases as the alignment between the observer and Player A increases, while Player A's punishment for winning decreases. These results appear to be consistent with Hypothesis 5.

Another pattern we observe in the data is that observers tend to punish both contestants heavily in the (relatively rare) case of breakdown. When neither player concedes in

[17] Subjects do not exhibit any learning effects due to punishment. The allocation of punishment by subjects is effectively equal before and after the first time subjects receives punishment themselves. Details can be found in the Appendix.

the Staring Contest, the observer deducts between 36.4 and 47.5 points from Player A and between 37.5 and 43.6 points from Player B. These deductions are roughly equal for both players and do not appear to depend on the alignment of interest between the observer and Player A. We did not anticipate this result in our theoretical analysis because the observer does not gain from punishing the contestants when there is breakdown; it is already the worst possible outcome and contestants should seek to avoid it.

One possible explanation for these behaviors is inequity aversion (Hatfield et al., 1978): Subjects dislike the resulting point distribution, and attempt to smooth it out by deducting points from the winning player. While our model does not account for it, caring about equality can be rational, and economic models of behavior often incorporate these preferences (Fehr and Schmidt, 1999; Goeree and Holt, 2000; Engelmann and Strobel, 2004).[18] This does not rule out instrumental uses of punishment such as deducting points from both sides in the event of breakdown. Rather, we suggest that deviations from our precise definition of rational play are themselves behaviorally coherent and interesting.

Table 4: **Average Deductions by Contest, Outcomes, and Treatment**

		Observer Ideal Point		
Contest Outcome		Obs 0	Obs 25	Obs 50
A Wins	Punishment to A	15.4	21.7	27
	Punishment to B	4.4	4.4	7.3
B Wins	Punishment to A	17.5	22.5	6.4
	Punishment to B	40	32.8	23
Breakdown	Punishment to A	36.4	47.5	37.5
	Punishment to B	43.6	43.5	37.5

Table 5 presents regression analyses of the observer's punishment behavior. In these

[18]Spite is not a useful alternative explanation for these results. As used in Levine (1998), spite is essentially the inverse of altruism. In our experiment, using spite as an explanation requires defining the *lack* of punishment as altruistic, and the use of punishment as spiteful. This does not explain, however, the finding that punishment is used more heavily against a specific player (B) when that player is the winner of the Staring Contest. A spiteful observer should use punishment roughly equally against all players that cost her points; discrimination between targets is not explained by a spite-base explanation.

models, we regress a measure of punishment on the actual stopping time, a set of dummy variables for the standoff outcomes, and a set of dummy variables for the observer's ideal point. The excluded category is the case where A wins and the observer's ideal point is $\theta_O = 50$. We used subject-level fixed effects to account for potential heterogeneity in observers' average deductions. The dependent variable for the first column is the total amount of punishment used by the observer. We find that total punishment is increasing in the duration of the Staring Contest, which is consistent with the notion that observers seek to encourage contestants to end the game early. More substantial differences in punishment behavior, consistent with the averages reported in Table 4, arise from differences in contest outcomes. Observers use the least amount of punishment when A wins the Staring Contest, use more total punishment when B wins, and use the most punishment when there is breakdown. We also find that observers use more punishment when $\theta_O = 25$ than when $\theta_O = 50$ but that there is no difference in total punishment when $\theta_O = 0$. Inequity aversion, then, is an aspect we see in the disparity of punishment used conditional on one side conceding. Punishment after a breakdown is roughly equal, which is still consistent with inequity aversion since the initial payoffs to both sides are equal (that is, punishment is not used to balance inequalities, but to scold players for hitting breakdown).

In the second and third columns of Table 5, we examine the effects of standoff outcomes and observer treatments on the points deducted from each player separately. Interestingly, we see that the observer deducts significantly fewer points from A when A loses than when A wins. This appears to be consistent with Hypothesis 4. However, we do not see any significant treatment effects of varying the observer's ideal point in column 2. Inequity aversion may play a role in this result: the observer may value a roughly equal point distribution for Players A and B more highly than she values any particular outcome of the contest. Different weightings on equity versus strictly rational play (as we define it) does not disprove our hypotheses — we cannot parse the competing theories given our data. In

Table 5: **Effects of Outcomes and Treatments on the Use of Punishment**

	(1) **Total Blame**	(2) **Blame to A**	(3) **Blame to B**	(4) **% of All Blame to A**
Actual Stopping Time	0.36**	0.04	0.32*	-0.00
	(0.13)	(0.12)	(0.12)	(0.00)
B Wins	17.07***	-7.37**	24.45***	-0.48***
	(3.04)	(2.79)	(2.80)	(0.05)
Breakdown	64.08***	25.82***	38.26***	-0.22*
	(7.21)	(6.62)	(6.62)	(0.11)
Obs 0	5.23	-1.07	6.30*	0.03
	(3.40)	(3.12)	(3.12)	(0.06)
Obs 25	9.12**	5.42	3.70	0.11*
	(3.40)	(3.12)	(3.13)	(0.06)
Constant	17.07***	19.39***	-2.32	0.79***
	(3.48)	(3.20)	(3.20)	(0.06)
R^2	0.13	0.05	0.20	0.38
R^2_w	0.29	0.10	0.26	0.40
R^2_b	0.00	0.01	0.06	0.31
N	396	396	396	223
N_g	66	66	66	57

Standard errors in parentheses.

Observer Ideal 50 and A Wins omitted as reference categories.

All models use subject-level fixed effects.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

column 3, we see that B is punished significantly more for winning than losing, which is also consistent with Hypothesis 4. And in this case, we also find that the amount of punishment is slightly greater when $\theta_O = 0$, which provides some support for Hypothesis 5.

The dependent variable in the fourth column is the percentage of total punishment used in the round applied to Player A. This captures the relative amount of blame assigned to Player A whenever punishment is used. These results are consistent with the results in columns 2 and 3. The intercept is 0.79, which means that Player A receives most of the punishment when A wins in the symmetric condition $\theta_O = 50$. But when Player B wins, the negative and statistically significant coefficient implies that punishment shifts from A to B—that is, to the winning player. Further, the coefficient for breakdown implies that the punishment is approximately equal. There is some evidence of an effect of varying the observer's ideal point, as the proportion of punishment applied to Player A increases when $\theta_O = 25$ relative to the symmetric case where $\theta_O = 50$, but it is in the wrong direction predicted by the theoretical analysis. Overall, the regression analysis implies support for Hypothesis 4. These choices in punishment allocation are, however, also consistent with a subject attempting to produce a more equitable final distribution of points.

4.3 Classification of Punishment Strategies

In this section, we provide additional insight into punishment behavior by classifying whether the observed allocations of blame are consistent with theoretical predictions or with other empirical regularities. We also classify subjects using these same categories by identifying subjects' modal punishment strategies. We learn that, while adherence to strict rationality (in terms of the magnitude of punishment) is low, most punishment follows the theoretical argument to punish the Staring Contest winner. This analysis provides further qualified support for our hypotheses concerning blame.

Following our theoretical analysis, we classify punishment as generally rational if the observer deducts points from the winner. More specifically, we say that punishment is *strongly rational* if it creates the strongest incentives for the game to end at $t = 0$: when the observer deducts 80 points or more from B when B wins (in any observer condition) or from A when A wins (but only if $\theta_O = 50$) and deducts 0 points from the loser. Weakening this requirement, we say that punishment is *weakly rational* if it targets the contestant according to the alignment of the observer's interest but less than the full magnitude required to induce a difference in contestants' behavior (i.e., less than 80 points). An even weaker version of this is for observers to *punish the winner* more than the loser under any condition (regardless of θ_0). These categories of rationality are also coded according to observer ideal point; all results can be read as pertaining to rational play that accounts for each particular value of θ_O.

We use two additional categories to describe allocations that also appear frequently in our data. Punishment is *equal* if the observer deducts an equal number of points from both players. We also classify rounds where *no punishment* is used at all.

We coded each instance of punishment with an indicator for the category that identified the allocation in that round. Next, we categorized each subject by finding their modal category. (Note that codings are not mutually exclusive at the level of observation since, for example, rational punishment is a subset of weakly rational punishment, nor are they mutually exclusive at the subject level in the case of ties.) This gives us a distribution of categories of play across rounds, and a distribution of general types of play across subjects. Table 6 presents these results.

We first note that subjects do not shy away from punishing the contestants. Non-zero punishment is used in over half of all periods. When punishment is non-zero, punishing the winner appears to be the prevailing strategy.[19] Although a small percentage of observations

[19]One indicative comment from subjects in the post-experiment questionnaire explicitly noted "I would deduct points from the player that did not concede. I was hoping this would make players more likely to concede in future rounds when I would be playing against them." Although this comment might suggest

Table 6: **Distribution of Play and Subjects by Category**

Behavior	Percent of Rounds ($N = 375$)	Percent of Subjects ($N = 66$)
Rational Punishment	11%	2%
Weakly Rational Punishment	26%	6%
Punish the Winner	37%	32%
Equal Punishment	9%	9%
No Punishment	46%	52%

fit the requirements of strongly rational play (11%), we do find that a quarter of observations can be classified as weakly rational (26%) and that even more rounds are consistent with punishing the winner (37%). When we condition on the use of non-zero punishment, 21% of such observations can be classified as strongly rational, nearly half (48%) count as weakly rational, and more than two-thirds (69%) are consistent with punishing the winner. We observe very few rounds with equal punishment (9%), while the observer uses no punishment whatsoever in slighly less than half of the rounds (46%). The subject-level classifications are generally consistent with the observation-level findings except that few subjects consistently use strongly rational punishment. This does not, indeed cannot, include explicit attempts at leveling the final point distribution in our definition of rationality. Overall, our classification analysis finds a high level of punishment targeted at the contest winner, which constitutes further qualitative support for our hypotheses.

4.4 Individual Differences in Behavior

At the end of the experiment, subjects provided their gender, ideology, party affiliation, and a description of their approach (if any) to playing the game.[20] We can use these data to

repeated play considerations, it is also consistent with subjects' understanding of learning and experience—that behavior adjusts to incentives over time. In contrast, the comments of subjects who did not use any punishment suggested they did so expecting reciprocity, for example: "I never deducted anyone's points, didn't want it to happen to me[.]"

[20]Subjects completed questionnaires via computer. They comprised a mixture of item selection and free-entry text fields. Note that six subjects did not complete the experiment and therefore did not fill out the

assess whether behavior in the Staring Contest varies with individual-level characteristics, and indeed, we find that differences are associated with gender and ideology. In general, we find evidence that women used more intermediate values for intended stopping times, that women use less punishment relative to men, and that being more conservative is associated with greater allocation of blame in general.

Recall that subjects entered an intended waiting time that had a small chance (1 in 10) of being implemented as their actual stopping time. These times are a form of commitment to conceding by a certain time in a non-trivial number of rounds, and thus are a significant strategic consideration for subjects. Plotting the distribution of these times by gender demonstrates a distinct difference in the use of stopping times. Figure 4 shows women enter extreme stopping times less frequently than men, providing times most often at intervals of five seconds.[21] The most striking difference is that men enter $t = 30$ for the intended stopping time over twice as often, and enter $t = 29$ almost as frequently as they enter $t = 30$. Note also that times within 5 seconds of each extreme (that is, $t \in [0, 5)$ and $t \in (25, 30]$) make up a larger proportion of the distribution for women than for men.[22]

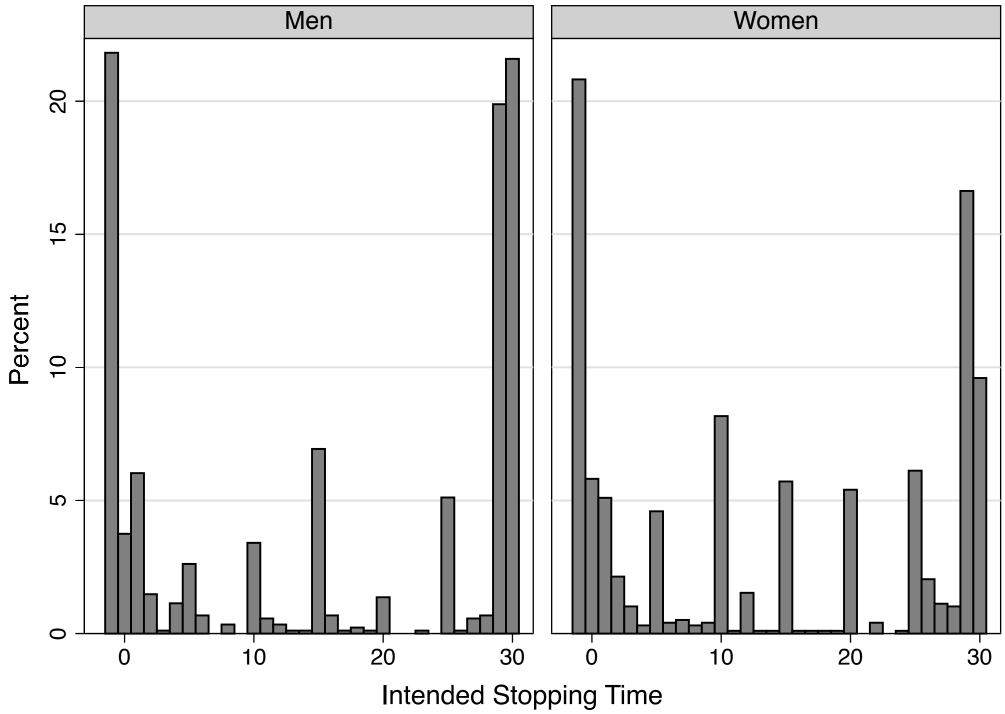
Women exhibit different strategies for play than men, allocating less punishment overall and deducting points in a more equitable manner. First, we show the gender-based differences in the use of punishment. Table 7 summarizes the deductions made by men and women, given the outcome of the contest and the observer ideal points. In line with Hypothesis 5, both men and women tend to punish B more for winning than A when the observer's ideal is 0 or 25. However, except for A winning when the observer's ideal is 25, women deduct considerably fewer points than men in the same conditions. Further, when

post-experiment questionnaire. Thus, there are 66 subjects for this section of the analysis.

[21] A Kolmogorov-Smirnov test rejects (at the $p < 0.001$ level) the null hypothesis that these distributions are equal.

[22] Regression analysis supports the presence of a gender-based difference. However, the negative and significant coefficient on the gender variable only shows that women enter lower intended stopping times on average, a fact already obvious in the graph.

Figure 4: **Distribution of Intended Stopping Times, by Gender**



the observer's ideal point is 0 or 50, men punish breakdown severely, while women punish breakdown either moderately (for $\theta = 0$) or not at all (for $\theta = 50$).

Table 7: **Average Deductions by Outcomes, Treatment, and Gender**

		Observer Ideal Point		
Contest Outcome		0	25	50
A Wins	Women	18.5	30.9	23.7
	Men	21.3	20.7	44.7
B Wins	Women	48.3	37.5	30.1
	Men	67	70.9	28.5
Breakdown	Women	53.3	94.3	0
	Men	100	83.3	100

Women appear more even-handed in their allocation of blame than men. Table 8 reprises Table 6 by categorizing play at the subject-level, disaggregated by gender. None of the women in our sample use strongly or weakly rational punishment, while fewer women punish the winner. Conversely, none of the men use equal punishment and women are more likely than men to refrain from using any punishment at all.[23] These findings support recent findings that men tend to play more (economically) optimal, or competitive, strategies in games of shared value and resource allocation than women, who are more likely to adopt egalitarian, or cooperative, strategies (Croson and Buchan, 1999; Kennelly and Fantino, 2007; Van den Assem, Van Dolder and Thaler, 2012). Furthermore, our results extend the reach of such findings to explicitly political scenarios.

Table 8: **Distribution of Rational Play by Gender**

Behavior	Men ($N = 32$)	Women ($N = 34$)
Rational Punishment	3%	0%
Weakly Rational Punishment	13%	0%
Punish the Winner	41%	24%
Equal Punishment	0%	18%
No Punishment	41%	62%

Subject ideology also appears to correlate with differences in the use of punishment. Table 9 shows that more conservative subjects punish more heavily. The distribution of ideological types is less even than for gender, with 23 liberal subjects, 29 moderate, and only 14 conservative, making the estimates for conservatives noisier. Differences in punishment are still notable. Biased observers increase their use of punishment as they become more ideologically conservative. When the observer's ideal point is either 0 or 25, punishment always increases under the same outcome: moderates punish more than liberals, conservatives punish more than moderates. However, for conservative, neutral observers, the trend reverses: they use less punishment than even liberals under the same outcome. In line with

[23] Some subjects might consider multiple rounds of the contest as a repeated game. If so, it is possible that either evenly distributing punishment or deducting no points at all could be a rational choice. Alternatively, the unwillingness to use punishment may stem from altruism, reciprocity, or other-regarding concerns.

Table 9: **Average Deductions By Outcomes, Treatment, and Ideology**

		Observer Ideal Point		
Contest Outcome		0	25	50
A Wins	Liberal	13.7	20.2	34
	Moderate	23.6	25.2	37.9
	Conservative	24.4	37.2	28.3
B Wins	Liberal	40.3	39.3	30
	Moderate	62.8	56	32.2
	Conservative	68.9	81.2	21.5
Breakdown	Liberal	30	100	66.7
	Moderate	100	87.5	100
	Conservative	100	80	—

rational play, Player B is punished more severely for winning than Player A. The only deviation is the conservative, neutral observer who punishes Player B less for winning than in any other treatment or outcome. Finally, breakdown is punished most severely by all observer ideologies, with the exception of liberal subjects when $\theta = 0$.

Examining rational play by ideology further supports our finding that subjects rarely engage in *strictly* rational play, but do frequently apportion most blame to the Staring Contest winner. Table 10 presents these results. Only liberal subjects can be classified as using strictly rational punishment, and then only rarely (4%). Liberals are also most likely to employ a weakly rational strategy (9%). However, conservative subjects are the most likely to punish the winner, almost doubling the frequency for liberals (43% vs 22%, respectively). Also, conservative observers are the most likely of the ideological types to dole out equal punishment to all players (14%), at half-again the rate of liberals (9%), and double that of moderates (7%).

Individual differences in play and punishment exist between men and women, and among ideological types. Women and men conceded at different rates, and responded differently when playing the same role; women concede roughly evenly as Player A or B, whereas

Table 10: **Distribution of Rational Play by Ideology**

Behavior	Liberals ($N = 23$)	Moderates ($N = 29$)	Conservatives ($N = 14$)
Rational Punishment	4%	0%	0%
Weakly Rational Punishment	9%	3%	7%
Punish the Winner	22%	34%	43%
Equal Punishment	9%	7%	14%
No Punishment	65%	41%	50%

men concede less as Player A. The presence of this effect is independent of subjects' strategic considerations, in contrast to work on gender in bargaining that generally makes gender known to all parties (Putnam and Jones, 1982; kan Holm and Engseld, 2005; Sutter et al., 2009; Cadsby, Servátka and Song, 2010). Further research on this effect would contribute to understanding differences that arise for intrinsic reasons rather than from strategic responses to gender. Ideology also matters in this study, with growing conservatism correlating with both greater use of punishment in general, and an unwillingness to concede as Player A. Similar to gender, ideology in bargaining experiments is usually modeled as common knowledge (Banas and Parks, 2002; Wade-Benzoni et al., 2002) but is important to understand as a personal trait affecting evaluations of politics.[24]

5 Discussion

In this paper, we examined how blame could be used as a mechanism for voters to control their representatives, focusing on how observers apportion blame and how the anticipation of blame affects political outcomes. We found that, in our experimental setting, observers use blame as punishment for delaying the resolution of a political standoff. Overall levels of punishment increase when observers' interests are more closely aligned with one side of the

[24]One interesting exception is Knight and Ensminger (1998), which considers personal ideology in the development of social norms, rather than through mutual evaluation of a bargaining partner's ideology.

fight, though no one is entirely spared and both sides are heavily punished for breakdowns. Thus, our findings suggest that interests alone do not determine the target of blame, cutting against the conventional wisdom that observers rarely blame copartisans for bad outcomes, as this behavior is broadly consistent with a rational strategy observers can use to induce better outcomes for themselves.

We also found that while blame does not significantly alter Staring Contest *outcomes*, it does induce representatives to reduce the *duration* of standoffs. The rational anticipatation of blame drives players to terminate contests earlier, thereby reducing waiting costs for all players. Shorter delays can be seen as a "bipartisan" improvement in our experiment as well as in political reality. But breakdown still occurs, of course. Our results suggest breakdown also arises when political elites fail to fully anticipate the consequences of protracted disagreements.

While the experimental evidence generally supports our hypotheses, inequity aversion may help explain deviations. Punishment is widely used, but is only mildly affected by observer ideal point treatments; subjects may care more about the final point allocation than their induced affiliation with Player A. Furthermore, "Equal Punishment" is a rarely used strategy for players of any gender or ideology. Deducting the same number of points from both sides does not satisfy a desire for equality of outcomes. A different structure for the Staring Contest would be necessary to fully tease out the prevalence of these incentives (and to separate them from risk aversion (Carlsson, Daruvala and Johansson-Stenman, 2005). The gap between our predictions of rational play and actual subject choices cannot be read simply as the inability to predict outcomes based on our model. Rather, we take these findings to suggest that subjects demonstrate behavioral regularities that could be, but in the current form are not, reflected in our model of political standoffs.

Extensions to our work would move beyond our focus on how voter blame impacts negotiating behavior during standoffs that are already underway. A single-shot game highlighted how an audience to the standoff decides to allot punishment at a single point in time. Of course, politics is frequently a repeated game, and expanding our experiment may add to the substantive findings reported here. Clearly, some subjects already considered the potential impact of punishment in one round on their outcomes in future rounds. In addition, pre-standoff communication might introduce interesting incentives, asking the observer to consider the possible avoidance of the standoff when allocating blame. This paper serves as a foundation for examining the important, and recurring, interaction between elected representatives and the expressed desire of voters during high-stakes political brinksmanship. In 2011, Standard & Poor's downgraded the U.S. Treasury partly because Congress' partisan showdowns repeatedly threatened default on loan service payments.[25] Prompt resolution of subsequent debt limit battles convinced another ratings firm to return the U.S. credit rating to full AAA status.[26] Similarly, in 1995, Speaker Newt Gingrich stood firm against President Bill Clinton over the federal budget, bolstered by the belief that the public would blame the president for a shutdown (Drew, 1997). Gingrich held out, and the government shut down for 27 days. While media labeled Gingrich the winner, the public clearly blamed Republicans more than Democrats.[27] Speaker John Boehner overestimated the public's support for his stance, refused to blink, and allowed the government to shut down again. Here too, despite Boehner "winning" by (temporarily) preventing an increase in the debt ceiling, the public laid the majority of blame at his feet.[28] Following these episodes, Gingrich and Boehner seemed to learn their lesson, as later battles ended in favor of their opponents.[29]

[25] Damian Paletta and Matt Phillips, "S&P Strips U.S. of Top Credit Rating", *Wall Street Journal*, 6 August, 2011.

[26] Jim Puzzanghera, "Fitch takes U.S. credit rating off downgrade watch after debt deal", *Los Angeles Times*, 21 March, 2014.

[27] CNN/USA/Gallup Poll, "Americans blame GOP for budget mess", *CNN*, 15 November 1995.

[28] Paul Steinhauser, "CNN Poll: GOP would bear the brunt of shutdown blame", *CNN Politics*, 30 September 2013.

[29] Paul Kane, Robert Costa, Ed O'Keefe, "House passes 'clean' debt-ceiling bill, ending two-week showdown", *Washington Post*, 11 February, 2014.

References

Abreu, Dilip and Faruk Gul. 2000. "Bargaining and Reputation." *Econometrica* 68(1):85–117.

Alcañiz, Isabella and Timothy Hellwig. 2011. "Who's to Blame? The Distribution of Responsibility in Developing Democracies." *British Journal of Political Science* 41(2):389–411.

Alicke, Mark D. 2000. "Culpable Control and the Psychology of Blame." *Psychological Bulletin* 126(4):556–574.

Banas, Joseph T and Judi M Parks. 2002. "Lambs Among Lions? The Impact of Ethical Ideology on Negotiation Behaviors and Outcomes." *International Negotiation* 7(2):235–260.

Bengtsson, Å sa. 2004. "Economic Voting: The Effect of Political Context, Volatility and Turnout on Voters' Assignment of Responsibility." *European Journal of Political Research* 43(5):749–767.

Brehm, John and Scott Gates. 1997. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, MI: University of Michigan Press.

Brewin, Chris R and David A Shapiro. 1984. "Beyond Locus of Control: Attribution of Responsibility for Positive and Negative Outcomes." *British Journal of Psychology* 75(1):43–49.

Brown, Adam R. 2010. "Are Governors Responsible for the State Economy? Partisanship, Blame, and Divided Federalism." *The Journal of Politics* 72(03):605–615.

Cadsby, C Bram, Maroš Servátka and Fei Song. 2010. "Gender and Generosity: Does Degree of Anonymity or Group Gender Composition Matter?" *Experimental Economics* 13(3):299–308.

Carlsson, Fredrik, Dinky Daruvala and Olof Johansson-Stenman. 2005. "Are People Inequality-Averse, or Just Risk-Averse?" *Economica* 72(287):375–396.

Celetani, Marco, Drew Fudenberg, David K Levine and Wolfgang Pesendorfer. 1996. "Maintaining a Reputation Against a Long-lived Opponent." *Econometrica: Journal of the Econometric Society* 64(3):691–704.

Chaudhuri, Ananish. 2011. "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature." *Experimental Economics* 14(1):47–83.

Croson, Rachel and Nancy Buchan. 1999. "Gender and Culture: International Experimental Evidence from Trust Games." *American Economic Review* pp. 386–391.

De Vries, Catherine E and Sara B Hobolt. 2012. "Do Voters Blame Governments for Social Spending Cuts? Evidence from a Natural Experiment.".

Drew, Elizabeth. 1997. *Showdown: The Struggle Between the Gingrich Congress and the Clinton White House*. Touchstone.

Embrey, Matthew, Guillaume R Fréchette and Steven F Lehrer. 2014. "Bargaining and Reputation: An Experiment on Bargaining in the Presence of Behavioural Types." *The Review of Economic Studies* pp. 0–29.

Engelmann, Dirk and Martin Strobel. 2004. "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments." *American Economic Review* pp. 857–869.

Fearon, J D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* pp. 577–592.

Fehr, Ernst and Klaus M Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114(3):817–868.

Fehr, Ernst and Klaus M Schmidt. 2000. "Fairness, Incentives, and Contractual Choices." *European Economic Review* 44(4-6):1057–1068.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic punishment in humans." *Nature* 415(6868):137–140.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10(2):171–178.
URL: *http://dx.doi.org/10.1007/s10683-006-9159-4*

Goeree, Jacob K and Charles A Holt. 2000. "Asymmetric Inequality Aversion and Noisy Behavior in Alternating-offer Bargaining Games." *European Economic Review* 44(4):1079–1089.

Groseclose, Tim and Nolan McCarty. 2001. "The Politics of Blame: Bargaining Before an Audience." *American Journal of Political Science* 45(1):100–119.

Guala, F. 2012. "Reciprocity: Weak or strong? What Punishment Experiments Do (and Do Not) Demonstrate." *Behavioral and Brain Sciences* 35(1):1.

Hatfield, Elaine, Elaine Hatfield Walster, G William Walster and Ellen Berscheid. 1978. *Equity: Theory and research*. Allyn & Bacon.

Hellwig, Timothy. 2010. "Elections and the Economy." *Comparing Democracies* 3:185–201.

Hood, Christopher. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton University Press.

Iyengar, Shanto. 1989. "How Citizens Think about National Issues: A Matter of Responsibility." *American Journal of Political Science* 33(4):878–900.

kan Holm, Håand Peter Engseld. 2005. "Choosing Bargaining Partners—An Experimental Study on the Impact of Information about Income, Status and Gender." *Experimental Economics* 8(3):183–216.

Kennelly, Arthur and Edmund Fantino. 2007. "The Sharing Game: Fairness in Resource Allocation as a Function of Incentive, Gender, and Recipient Types." *Judgment and Decision Making* 2(3):204–216.

Knight, Jack and Jean Ensminger. 1998. *Conflict Over Changing Social Norms: Bargaining, Ideology, and Enforcement.* Stanford, CA: Stanford University Press.

Kousser, Thad and Justin H Phillips. 2009. "Who blinks first? legislative patience and bargaining with governors." *Legislative Studies Quarterly* 34(1):55–86.

Kovenock, Dan and Brian Roberson. 2009. "Is the 50-state strategy optimal?" *Journal of Theoretical Politics* 21(2):213–236.

Kreps, David M, Paul Milgrom, John Roberts and Robert Wilson. 1982. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma." *Journal of Economic Theory* 27(2):245–252.

Kreps, David M and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27(2):253–279.

Lagnado, David A and Shelley Channon. 2008. "Judgments of Cause and Blame: the Effects of Intentionality and Foreseeability." *Cognition* 108(3):754–770.

Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiment." *Review of Economic Dynamics* 1(3):593–622.

Lewis-Beck, Michael S and Martin Paldam. 2000. "Economic Voting: An Introduction." *Electoral Studies* 19(2):113–121.

Lewis-Beck, Michael S and Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3(1):183–219.

Lipsky, Michael. 1980. *Street-Level Bureaucracy.* Russell Sage Foundation.

Malhotra, Neil and Alexander G Kuo. 2007. "Attributing Blame: The Public's Response to Hurricane Katrina." *The Journal of Politics* 70(01):120–135.

Masclet, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval. 2003. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review* 93(1):366–380.

Paldam, Martin. 2008. Vote and Popularity Functions. In *Readings in Public Choice and Constitutional Political Economy*, ed. Charles Kershaw Rowley and Friedrich Schneider. New York: Springer pp. 533–550.

Peffley, Mark. 1984. "The Voter as Juror: Attributing Responsibility forEconomic Conditions." *Political Behavior* 6(3):275–294.

Putnam, Linda L and Tricia S Jones. 1982. "Reciprocity in Negotiations: An Analysis of Bargaining Interaction." *Communication Monographs* 49(3):171–191.

Roth, Alvin E, ed. 1985. *Game-Theoretic Models of Bargaining*. Cambridge: Cambridge Univ. Press.

Schultz, K A. 2001. "Looking for audience costs." *Journal of Conflict Resolution* 45(1):32–60.

Shaver, Kelly G. 1985. *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. Springer-Verlag New York.

Sutter, Matthias, Ronald Bosman, Martin G Kocher and Frans van Winden. 2009. "Gender Pairing and Bargaining—Beware the Same Sex!" *Experimental Economics* 12(3):318–331.

Tomz, Michael. 2007*a*. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61(4):821.

Tomz, Michael. 2007*b*. *Reputation and International Cooperation: Soverign Debt across Three Centuries*. Princeton, NJ: Princeton University Press.

Trachtenberg, Marc. 2012. "Audience Costs: An Historical Analysis." *Security Studies* 21(1):3–42.

Treier, Shawn and D Sunshine Hillygus. 2009. "The nature of political ideology in the contemporary electorate." *Public Opinion Quarterly* 73(4):679–703.

Van De Ven, Jeroen and Marie Claire Villeval. 2014. "Dishonesty under scrutiny." *Journal of the Economic Science Association* pp. 1–14.

Van den Assem, Martijn J., Dennie Van Dolder and Richard H. Thaler. 2012. "Split or Steal? Cooperative Behavior when the Stakes are Large." *Management Science* 58(1):2–20.

Wade-Benzoni, Kimberly A, Andrew J Hoffman, Leigh L Thompson, Don A Moore, James J Gillespie and Max H Bazerman. 2002. "Barriers to Resolution in Ideologically Based Negotiations: The Role of Values and Institutions." *Academy of Management Review* 27(1):41–57.

Wawro, Gregory J and Eric Schickler. 2006. *Filibuster: Obstruction and lawmaking in the US Senate*. Cambridge Univ Press.

Weale, Albert. 2002. *Risk, Democratic Citizenship and Public Policy.* Vol. 2 Oxford: Oxford University Press.

Weaver, R Kent. 1986. "The Politics of Blame Avoidance." *Journal of Public Policy* 6(4):371–398.

Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1):35–64.

Weiner, Bernard. 1995. *Judgments of Responsibility: A Foundation For A Theory of Social Conduct.* New York: Guilford Press.

Appendix A Learning Effects

Subjects may learn to play the Staring Game differently during the course of the experiment. If this occurs, and it affects the player's choices in ways that undermine our hypotheses, we must correct for them in the design of the experiment. In the case of time-based effects, we might be concerned that players may alter their strategy as they improve. Earlier rounds would then represent a different level of information than later rounds, since players have deduced the "correct" way to play the staring game. The second issue, punishment-based learning effects, suggests subjects play in one fashion until they experience punishment themselves, then play differently afterwards. The subjects learn, in other words, to be more punitive only when they have had points deducted. In this appendix we address two sources of concern – time-based effects and punishment-based effects – and show that they are not a problem.

A.1 Time-Based Learning Effects

Subjects that choose to concede do so more frequently during the last five seconds of the Staring Game as the number of rounds increases. While they appear to understand the basic game at the outset, after the first rounds they prefer waiting to concede near the end to conceding early.

Table 11 shows several probit models of this behavior. For each model, a player's choice to concede was categorized by either the exact or approximate time of concession. In the first column, subjects were classified as stopping at "Any Extreme" if their choice to concede was at $t < 5$ or $t > 25$. As noted in the main text, looking at time beyond the very first and last second accounts for differing subject response times. The significant coefficient shows that there is an increased probability to concede at one extreme or another as the number of periods increases. The next models examine whether the concessions are happening at one or both of these extremes. Columns 2 and 3 captures those rounds where players concede either within the first five seconds ($t = 0...4$) or just the first second, respectively. Clearly, concession at these times is not happening more frequently in later rounds. Columns 4 and 5 perform a similar analysis for stopping in the last five seconds ($t = 26...30$) and the last second, respectively. As the number of periods grows, players are more likely to concede in the last five seconds, though not at the very last second.

The presence of a time-based learning effect does not alter our substantive findings. Our treatment – the introduction of observers – occurs after a number of rounds are played, so that any effects present are constant for a player and stable during treatment. In addition, subjects learn to play according to the predictions of the Staring Game model.

A.2 Punishment-Based Learning Effects

Another possible concern takes this form: do subjects change their behavior *after they experience punishment*? In terms of our experiment, this translates to: do subjects adopt one

Table 11: **Time-Based Learning Effects**

	(1) Any Extreme	(2) First 5 Seconds	(3) First Second	(4) Last 5 Seconds	(5) Last Second
Period	0.11***	0.03	0.05	0.09**	-0.06
	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)
Constant	-0.31	-1.01***	-2.41***	-0.94***	1.99***
	(0.18)	(0.23)	(0.44)	(0.21)	(0.31)
$\ln(\sigma_u^2)$	-1.49**	-0.53	0.11	-0.92*	-1.07
	(0.49)	(0.37)	(0.52)	(0.38)	(0.68)
N (rounds)	398	398	398	398	398
N (subjects)	72	72	72	72	72

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

strategy in all rounds before they have points deducted by an observer, and another strategy in all rounds after?

Concerns about punishment-based effects are not warranted for this experiment. To show this, we categorized each round for each subject as being before or after the first round in which they received a deduction, then compared the use of punishment within subjects. Table 12 shows the average deductions taken by observers, first in those rounds before they received any punishment themselves, then after. T-tests show Punishment use is statistically equal in these two conditions, and across the targets of punishment ($p < 0.2$ for Player A, and $p < 0.9$ for Player B).

Table 12: **Punishment Use Before and After Receiving First Deductions**

	Mean Punishment to A	**Mean Punishment to B**
Pre-Punishment	16.7	17.5
	(28.1)	(28.8)
Post-Punishment	21.2	19.9
	(32.8)	(32.8)

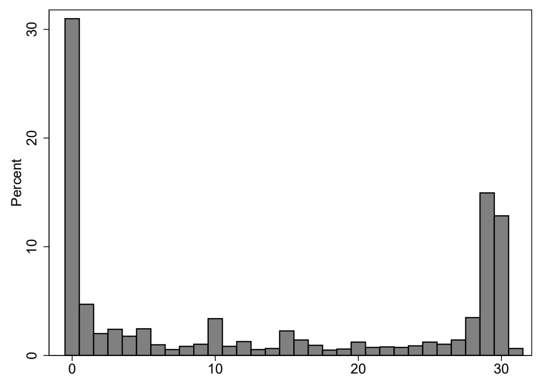
Standard deviations in parentheses.

Appendix B Stopping Time Details

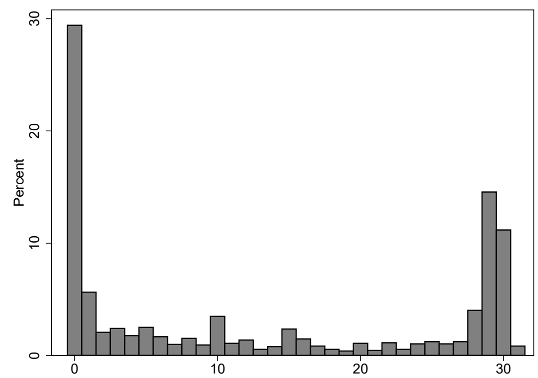
We provide here further details about the relationship between stopping times and player roles (Players A and B), and the relationship between stopping times and punishment.

Figure 5 shows the distributions of stopping times for Players A and B (the only roles that can end the Staring Contest). The clear similarity indicates that there is no substantive difference in the stopping times between Players A and B.

Figure 5: **Stopping Times for Contestant Roles (Players A and B)**



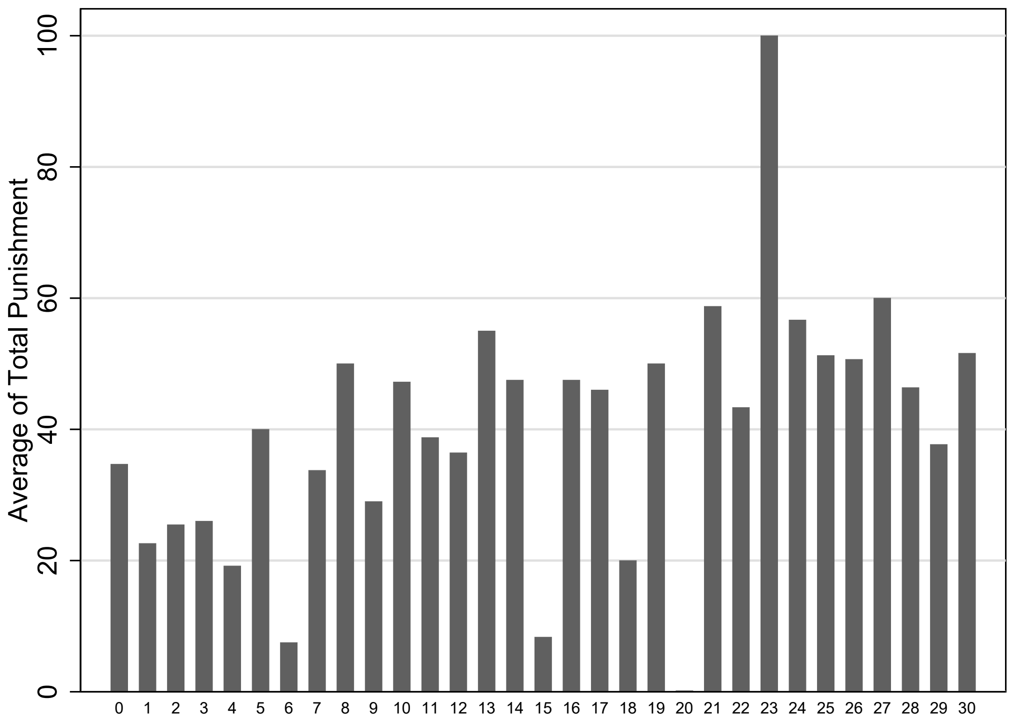
(a) **Player A**



(b) **Player B**

Figure 6 shows the distribution of average punishment used by time the Staring Contest was stopped. As demonstrated in more granular detail in the main text of the paper, the amount of punishment used is an increasing function of the time the Staring Contest lasts. Interestingly, there is a spike at $t = 23$, showing that, on average, all available punishment was used when the Staring Contest was ended at this point. Of course, this may be accounted for by the fact that the contest rarely ended at $t = 23$, so the number of times punishment could be allocated might be as little as one.

Figure 6: **Average Punishment by Stopping Time**



Instructions

General Information

This is an experiment on the economics of bargaining. The University of Pittsburgh has provided funds for this research.

The experiment consists of two parts, and each part is divided into a number of rounds. During each round, you will earn points, and points will be converted to cash at the rate of $1 for every 20 points, rounded to the nearest quarter. At the end of the experiment, we will randomly select **one round** to count for your payment. Follow the instructions closely, as we will explain how you will earn money and how your earnings will depend on the choices that you make. In addition to the $5 participation payment, these earnings will be paid to you, in cash, at the end of the experiment.

You will be paid your earnings privately, meaning that no other participant will find out how much you earn. Each participant has a printed copy of these instructions and you may refer to them at any time during the experiment.

If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Also, please ensure that your cell phones are turned off and put away. Participants intentionally violating the rules will be asked to leave and may not be paid.

Part I

In Part I of the experiment, you will play 12 rounds of a Staring Contest Game. In each round, you will be randomly matched with one other participant. You will not know the identity of the other participant you are matched with, and they will not know your identity.

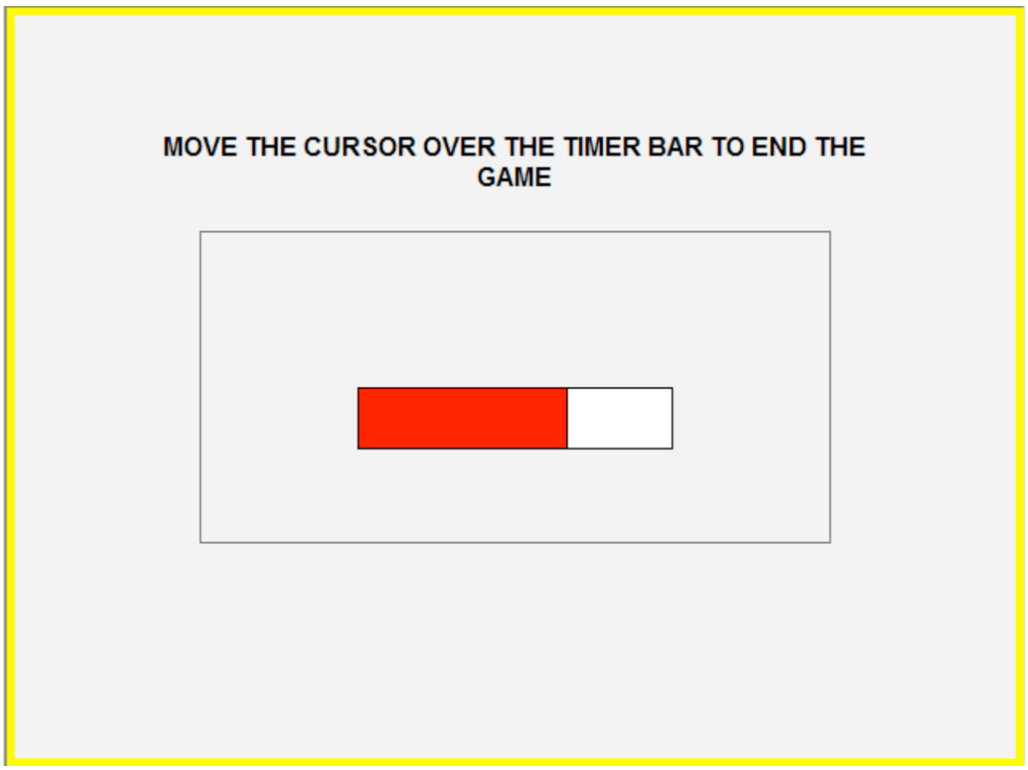
The Staring Contest Game

In every round, you will be either Player A or Player B, and each player is assigned a **target number**. Player A's target number is 0 and Player B's target number is 100.

Every game has two possible **outcome numbers**, either A = 10 or B = 90, and the outcome number is determined by the result of the Staring Contest Game. Notice that outcome A is closer to Player A's target and outcome B is closer to Player B's target.

The rules of the game are simple. At the beginning of the round, a clock will begin ticking down. It will tick down for 30 seconds in 1-second intervals. Your only decision during the game is deciding **when to concede** by stopping the clock. If you stop the clock, then the outcome number of the game is the one closer to the other player's target. That is, if Player A concedes, the outcome number is B; if Player B concedes, the outcome number is A.

Below is an example of the Staring Contest screen. The red bar will decrease every second, until it is gone at 30 seconds. To concede and stop the contest, simply **move your cursor over the bar**. Do not click the mouse**.** As soon as you move the cursor over the bar, the game will end. Moving the cursor anywhere outside timer bar will not affect the countdown.



Payoffs

Your earnings in each round depend on your target number, the outcome number, and the time at which the clock is stopped. Specifically, you will earn 350 points minus the absolute difference between the outcome number and your target. We will also subtract 1 point for each second that the clock ticks down before it is stopped. However, if neither player stops the clock, then your payoff will be 190. Thus, if one of the players stops the clock, then your payoff is given by the following formula:

$$\text{Payoff} = 350 - |\text{Outcome} - \text{Target}| - \text{Seconds passed}$$

For example, if Player A stops the clock at 10 seconds, then Player A's payoff is 350 - |90 – 0| – 10 = 250, while Player B's payoff is 350 - |90 – 100| - 10 = 330. If Player B instead stops the clock at 25 seconds, then Player A's payoff is 350 - |10 – 0| - 25 = 315 while Player B's payoff is 350 - |10 – 100| - 25 = 235. Of course, these are just a few examples.

Note that you do not have to keep track of your earnings, which will be recorded by the computer.

Intended Stopping Time

Before every round, we will also ask you for an intended stopping time. You should enter a number indicating the **number of seconds between 0 and 30 you plan to wait before stopping**. For instance, if you want to concede after 10 seconds have passed, you should enter 10 in the box. If you want to wait until 25 seconds have passed, you should enter 25 in the box.

The computer will then randomly draw a number from 1 to 10 before the round begins to determine if we implement your intended stopping time. We will draw the random numbers separately and independently for each participant in each round. If the random number is 1, then we will implement your intended stopping time as your actual stopping time. If the random number is any number from 2 to 10, then you will play the Staring Contest Game exactly as we described it above. Thus, there will be a small (1 in 10) chance that this intended stopping time will be your actual stopping time.

Part II

Part II consists of 18 rounds, and in every round you will be randomly assigned to a group of three participants. As before, you will not know the identities of the other members of your group, and they will not know yours.

In every group of three, there are two **contestants** and one **observer**. You will play a total of 12 rounds as a contestant and 6 rounds as an observer. The contestants will play the Staring Contest Game exactly as we described it in Part I. The observer's role is described next.

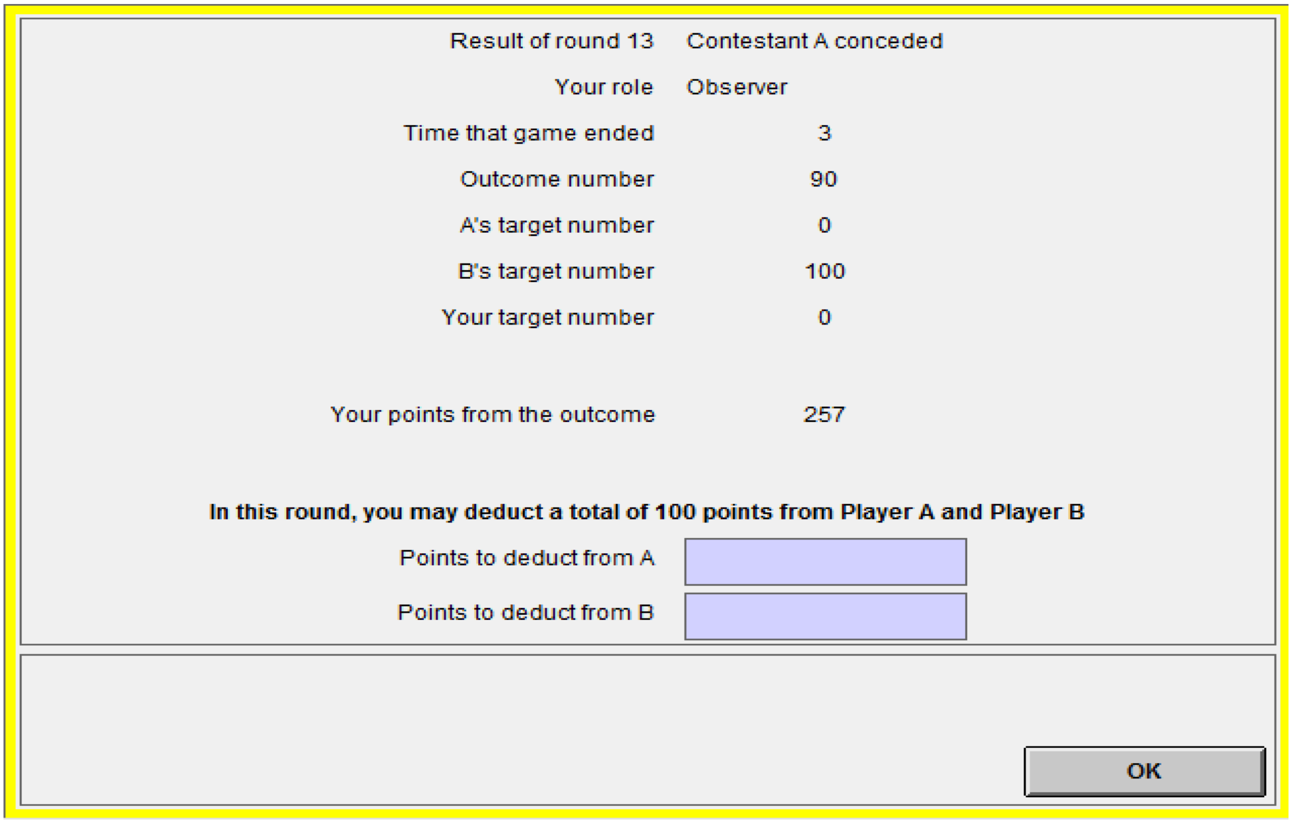
Observers

Like the contestants, the observers will also have target numbers. These target numbers will be different from the contestants' target numbers and will vary from round to round. These target numbers will be either 0, 25, or 50 and will be announced to all players at the beginning of the round.

During the Staring Contest, the observer has no decisions to make. Observers will see the timer count down and wait for the contestants to make their decisions. When the Staring Contest is finished, observers will see the results and can choose to **deduct** points from **one or both** of the contestants. The observer can choose any number of points to deduct from each of the contestants (including zero) provided that the total number deducted may not exceed **100 points**.

Note that the observer can choose to deduct some, all, or none of these points from Player A and/or Player B. The exact allocation of points to deduct is entirely up to the observer.

Below is an example of the screen observers will see at the end of the Staring Contest. You can see the areas where you may enter points to deduct. If you do not want to deduct points from Player A, B, or either, **you must enter zero** in the field. **You cannot leave them blank.**



Payoffs

As in Part I, earnings in each round of Part II depend on a player's target number, the outcome number, and the time at which the clock stopped. You will earn 350 points minus the absolute difference between the outcome number and your target, less 1 point for each second the clock ticks down before it is stopped. If neither contestant stops the clock, all players including the observer, receive a payoff of 190. We will also reduce each contestant's payoffs by the amount chosen by the observer.

If one player stops the clock in the Staring Contest, then contestants' payoffs are given by the following formula:

$$\text{Payoff} = 350 - |\text{Outcome} - \text{Target}| - \text{Seconds passed} - \text{Observer deduction}$$

while the observer's payoff formula is:

$$\text{Payoff} = 350 - |\text{Outcome} - \text{Target}| - \text{Seconds passed}$$

For example, suppose the observer's target is 25 and Player A stops the clock at 15 seconds so that the outcome number is 90. Suppose also that the observer deducts 40 points from Player A and 60 points from Player B. In this case, Player A's payoff is 350 – |90 - 0| – 15 – 40 = 205; Player B's payoff is 350 - |90 – 100| – 15 – 60 = 265; the observer's payoff is 350 - |90 – 25| = 285. Of course, this is just one of many possible outcomes of a round.