Time and Punishment: 
Blame and Concession in Political Standoffs*

Ian Palmer Cook† 
Jonathan Woon‡

January 2, 2017

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 the observer shortens the duration of standoffs and leads to outcomes that favor the observer. We also find that observers tend to punish the winning player and that the level of punishment depends on the alignment between the observer and players. The experimental data are qualitatively consistent with instrumental punishment, but the magnitude tends to be less than optimal, possibly reflecting subjects’ behavioral or emotional responses. (Keywords: Experiment; Behavioral Political Economy; War of Attrition; Punishment)

JEL Classifications: C78; D72; D74

*We thank Catherine Hafer, Kristin Kanthak, George Krause, Victoria Shineman, the guest associate editor, and two anonymous reviewers for helpful comments and discussion. A prior version was presented at the 7th Annual NYU-CESS Experimental Political Science Conference.
†ipc4@pitt.edu
‡Associate Professor, Department of Political Science, Department of Economics (secondary) and Pittsburgh Experimental Economics Laboratory, University of Pittsburgh. woon@pitt.edu (corresponding author)
“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
or give-and-take processes usually associated with deal-making and compromise.¹ Consider
recent struggles over U.S. fiscal policy. In September 2013, congressional Republicans de-
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 capit-
ulated.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.

¹We generally avoid the term “bargaining” throughout the paper. While standoffs occur during a bar-
gaining 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.
Scott Clement, “Republicans are losing the shutdown blame game”, The Fix, 4 October, 2013.
³John Breshnahan, Manu Raju, Jake Sherman, and Carie Budoff Brown, “Anatomy of a shutdown”,
Politico, 18 October, 2013.
⁴Paul Kane, Robert Costa, Ed O’Keefe, “House passes ’clean’ debt-ceiling bill, ending two-week show-
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 waiting is costly and the only choice is when to back down. This setting allows us to isolate the incentives for allocating blame for the outcome from other motives about discounting or reputation effects. It is also important for our purposes because the environment is one in which 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. 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 not only in the example above but also by political scientists studying subnational government (Kousser and Phillips, 2009) and international relations (Fearon, 1994).

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 instrumental 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. We then discuss our empirical results and conclude with a discussion of these findings and their general implications.
Our experiment yields several key results regarding blame and political standoffs. First, the presence of an observer decreases the duration of the Staring Contest, which is consistent with the fact that we find the audience tends to blame winners more than losers, especially when they are indifferent about who wins. Second, as the alignment between the audience and one of the standoff contestants increases, the contestant aligned with the observer wins the Staring Contest more often. This is consistent with our empirical finding that the audience punishes the non-aligned contestant more heavily for winning than the aligned contestant in these cases. While these patterns of punishment are qualitatively consistent with our definition of blame and our theoretical expectations, we also find that the magnitude of punishment is less than optimal, suggesting that blame may be emotional or behavioral. Furthermore, although punishing the winner might seem to be consistent with inequity aversion, such an alternative explanation cannot (fully) explain our finding that an audience punishes one contestant more heavily than another or why both contestants would be punished equally in the case of disagreement.

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 decision makers 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, 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). 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 oppo-

---

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.
nent’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. In their model, voters are the audience and blame takes the form of lowering approval ratings for the president if he vetoes a bill. However, their voters can only assign full blame to one side of what is a strategic interaction. In contrast, our experiment does not constrain how blame is assigned: the audience can blame neither, one, or both parties according to their assessment of both actors’ choices.

Although we motivated our experimental setting with high profile examples of budgetary standoffs at the national level between the United States Congress and the president, political standoffs are common in other domains (Wawro and Schickler, 2006) as well as at the subnational and international levels. For example, budgetary impasses in state governments are also modeled as wars of attrition (Alesina and Drazen, 1991; Kousser and Phillips, 2009; Kovenock and Roberson, 2009). The war of attrition is also central to models of international crises and the vast international relations literature on audience costs, a phenomenon in which domestic audiences can impose costs on political leaders and thereby affect the outcomes of crises (Fearon, 1994; Tomz, 2007b,a; Weeks, 2008). We complement this by making the cost mechanism—a subject of some debate (Schultz, 2001; Trachtenberg, 2012)—explicit.

---

6Our war of attrition framework is not one in which gridlock arises due to a Pareto-efficient status quo policy, such as for an “interior” status quo policy in a spatial model of lawmaking (e.g., Krehbiel (1998)). Nevertheless, the assumption that disagreement is “bad” for both sides does correspond to situations in such models where the status quo is “extreme” or far outside the “gridlock interval” (e.g., see Woon and Cook (2015)). Furthermore, political scientists studying American politics document that historically high levels of ideological polarization mean that parties and lawmakers are increasingly less likely to compromise, suggesting that the war of attrition is more relevant today than ever before.
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.\footnote{The general public goods literature is expansive. For a useful summary of punishment and cooperation in experimental settings, see Chaudhuri (2011).} 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

We designed a simple experiment that contrasts standoffs with and without an audience to investigate how the presence of an audience and the potential for punishment affects behavior during a standoff. In addition, our setting allows us to examine how an audience might actually punish the standoff contestants and to see how outcomes and punishment varies with the degree of alignment between the observer and the players. The latter objective is achieved by modeling outcomes as spatial policies and varying the ideal point of the observer.

In the baseline No Observer condition, 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 graphical 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. (When we introduce an observer, the spatial description allows us
to manipulate the alignment of preferences between the observer and the two contestants.) 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 insofar 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 \( \phi \) denotes disagreement, and the game ends at time \( t \) (in seconds), then player \( i \)'s payoff (denominated in points) is given by the function

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

where \( \theta_i \) 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 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 the Observer conditions, randomly matched groups of three participants played the Staring Contest Game with an audience for a total of 18 rounds. We designated two players\(^8\)

---

\(^8\)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.
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 the Observer condition 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 play the Staring Contest Game exactly as they do in the No Observer condition. 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.\footnote{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.} 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 the Observer condition, 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 A than Player B. And when $\theta_O = 0$, the observer is completely aligned with Player A.\footnote{Note that the observer can make conceding the worst possible outcome by deducting 80 points or more from the conceding player.} This alignment models partisan bias in the audience, while doing so only for one side keeps 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.

The experiment utilized a within-subject design, and we divided the total of 30 rounds into two parts (a block of 12 rounds of the No Observer condition and a block of 18 rounds of the Observer conditions). To guard against the possibility of order effects, we ran 5 sessions (72 subjects) with the No Observer condition as Part 1 (and Observer conditions as Part 2) and 3 sessions (42 subjects) with the order of the conditions reversed with the Observer conditions as Part 1 (and the No Observer conditions as Part 2).\textsuperscript{11}

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.

The experiments were conducted at Pittsburgh Experimental Economics Laboratory at the University of Pittsburgh. Subjects were recruited from the lab’s general subject pool and gave informed consent according to standard procedures. We read the instructions out loud to induce public knowledge and checked subjects’ comprehension using a multiple-choice quiz. At the conclusion of the experiment, subjects were paid privately in cash. We converted points to cash at the rate of $1.00 per 20 points, and the average payment was $18.75 (including a $5.00 show-up fee).

\textsuperscript{11}One subject voluntarily left the experiment during the last rounds of Part 1 in an Observer-first session. This forced us to involuntarily dismiss additional subjects to maintain an appropriate number of subjects per group. We use data from this session in the following results. Thus, the total number of subjects for rounds 1-10 is 114 while the number of subjects in rounds 11-30 is reduced to 108. The basic results do not change if these data are excluded.
3 Theoretical Expectations

The purpose of our theoretical analysis is to generate a reasonable set of behavioral expectations. Although we rely on payoff-maximization to guide our analysis, we emphasize that our goal is not to test any particular equilibrium theory. Indeed, as we note below, the game has multiple equilibria, both with and without an observer, so equilibrium theory has only weak predictive power. Rather, the principal aim of our experiment is to examine human behavior in a controlled setting. Thus, in our analysis we are primarily interested in 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 Staring Contest 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 \( t_A > t_B \). (This point, and the following analysis, is symmetric for Players A and 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 (because \( 340 - t_B > 260 - t_A \)).

Given that each player has an incentive to always outlast the other, the game has two asymmetric pure strategy Nash equilibria. In each equilibrium, one of the players never concedes while the other concedes immediately. The rationale is straightforward: if player \( j \) never concedes \( (t_j > 30) \), then it is a best response for player \( i \) to concede immediately \( (t_i = 0) \). There is no other equilibrium in which any player concedes at \( t_i > 0 \). Hence, if both
players are fully rational and correctly anticipate each others’ behavior, we should observe that Staring Contests always end immediately.

Behaviorally, however, the asymmetric Nash equilibrium outcomes involving immediate concession require coordination between the players that is unlikely to arise in a symmetric, dynamic environment. We can formulate alternative behavioral expectations based on the intuition that players may understand their incentives to wait but fail to recognize the equilibrium outcome (a form of incomplete strategic reasoning) or fail to coordinate on which player is to concede. If players are thus boundedly rational, then we might expect them to be willing to wait until the last second (attempting to outlast their opponent) but then unwilling to risk the Pareto inferior disagreement outcome. If both players wait until the last second, the tie-breaking rule implies an equal likelihood of obtaining a payoff of 310 (from winning) and a payoff of 230 (from losing) for an expected payoff of 270. Note that this outcome is preferable to, at best, conceding immediately and obtaining 260 points. Thus, if each player has an out-of-equilibrium expectation that the other player is willing to concede, but only at the last possible moment, then both players are willing to wait until the very end of the Staring Contest and the game should end at $t = 30$.

We can summarize the expectations of extreme stopping times, based on equilibrium analysis as well as our intuition about boundedly rational behavior, as the first hypothesis.

**Hypothesis 1** *In the Staring Contest Game without an observer, the game will either end immediately (if both players are fully rational) or both players will wait until the last moment before conceding (if both players are boundedly rational). We expect to observe waiting times at the extremes of 0 and 30 seconds.*

Now consider the game in the Observer condition. Since 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. That is, any allocation of punishment can be supported in equilibrium, so equilibrium theory does not generate sharp predictions regarding observer behavior.
Instead, we ask in our analysis how the observer might use punishment instrumentally to improve her payoffs. That is, since the observer is not indifferent between all possible outcomes (equilibria), how could a savvy payoff-maximizing observer design a punishment strategy that, if rationally anticipated by the contestants, induces the best possible Staring Contest outcome for herself? In the spirit of backward induction, this allows us to generate expectations about what the observer would do and then derive implications for how contestants would change their behavior in anticipation of the observer’s response. In our analysis, we concentrate on the equilibrium supported by the minimal amount of punishment that induces the best outcome for the observer.

When the observer’s ideal point is \( \theta_O = 50 \), she is indifferent between the outcomes \( a = 10 \) and \( b = 90 \) and therefore the best outcome is the one that minimizes the duration of the Staring Contest. That is, she does not care which contestant wins as long as \( t_A = 0 \) or \( t_B = 0 \). Indeed, 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 \( \theta_O = 0 \), the best outcome for the observer is for Player A to win (that is, for Player B to concede) the Staring Contest at \( t = 0 \). This yields the outcome \( a = 10 \). To achieve this, the observer can encourage 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

\(^{12}\)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 \( t_j = 0 \).
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 instrumentally rational 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, 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, the Staring Contest is more likely to end immediately, and disagreement is less likely 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 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.
We provided a set of theoretical expectations based on a baseline assumption of instrumental, selfish behavior. It is worth considering what alternative motivations for behavior might imply. For example, it is common to find that in many laboratory experiments subjects appear to act out of pro-social motivations such as fairness or inequity aversion (Fehr and Schmidt, 1999). In our setting, inequity aversion would imply punishing the winner (and only the winner) of the Staring Contest in order to reduce the inequality between the contestants. Thus, while Hypothesis 4 is consistent with inequity aversion, reducing inequality between contestants is inconsistent with Hypotheses 2, 3, or 5.

Another possibility is that subjects exhibit limited strategic sophistication, as in a level-k or cognitive hierarchy model of behavior (Camerer, Ho and Chong, 2004; Carpenter, Graham and Wolf, 2013; Shapiro, Shi and Zillante, 2014). More specifically, behavior consistent with Hypotheses 4 and 5 could result from a kind of level-0 emotional response or a behavioral form of blame consistent with expression of “displeasure” about the outcome of the Staring Contest. If the contestants are also level-0 and fail to anticipate how observers respond to their Staring Contest behavior, then we would not expect to find evidence for Hypotheses 2 or 3. On the other hand, if subjects are sufficiently sophisticated (level-1 or higher) in the sense that they understand and anticipate the observer’s punishment strategy (which would make sense given that subjects play both contestant and observer roles), then their behavior would be consistent with instrumental rationality and we would expect to see evidence for Hypotheses 2-5.

4 Findings

We first show how the presence of an Observer affects the duration and outcomes of the Staring Contest. Next, we analyze the patterns of punishment Observers use to produce these changes and then characterize the distribution of punishment strategies we observe. In
our analysis, we pool the data for all of our sessions regardless of the order of the Observer and No Observer conditions, as we find no evidence that the order affects our results (see Appendix A).

### 4.1 Contest Duration and Outcomes

Figure 1 shows the distribution of actual and intended concession times in the baseline Staring Contest without an observer. Actual stopping times clearly follow a bimodal distribution, with the majority of contests ending around $t = 0$ and near $t = 30$. These results are generally consistent with the expectations stated in Hypothesis 1. The game ends immediately (at $t = 0$) in 18% of rounds played and at the last possible second (at $t = 30$) in 16% of rounds played. If we allow for some error in waiting times, we find that 36% of rounds played end within the first 5 seconds, and another 36% end within the last 5 seconds. Intended waiting times are roughly similar. Nine percent of subjects indicated they intended to concede immediately (at $t = 0$), while 14% indicated they intended to wait until the very end (at $t = 30$). Further, 45% of all intended stopping times fall within the first and last five seconds. Thus, without an observer, the majority of games ended—or subjects stated their willingness to end the game—at the very beginning or end of the waiting period, possibly reflecting a mixture of equilibrium and boundedly rational behavior.

Introducing an Observer to the Staring Contest materially affects the duration of the game. Table 1 presents the average actual and intended stopping times by observer condition. While the difference between an Observer with an ideal point of 50 (fully neutral) and the

---

13 The Pearson correlation coefficient for actual and intended stopping times is $\rho = 0.25$. This is not an especially high correlation, so we present results for both actual and intended stopping times in this section. Nevertheless, the results for both measures 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 exactly $t = 0$. 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 B.
Figure 1: **Actual Stopping Times Bimodal, Intended Times Cluster**

(a) Actual Stopping Times

(b) Intended Stopping Times

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

<table>
<thead>
<tr>
<th></th>
<th>Actual</th>
<th>Intended Player A</th>
<th>Intended Player B</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Observer</td>
<td>Mean</td>
<td>15.4</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>621</td>
<td>621</td>
</tr>
<tr>
<td>Observer 50</td>
<td>Mean</td>
<td>14.3</td>
<td>17.8</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>207</td>
<td>207</td>
</tr>
<tr>
<td>Observer 25</td>
<td>Mean</td>
<td>13.5</td>
<td>18.1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>202</td>
<td>202</td>
</tr>
<tr>
<td>Observer 0</td>
<td>Mean</td>
<td>12.7</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>204</td>
<td>204</td>
</tr>
<tr>
<td>Total</td>
<td>Mean</td>
<td>14.5</td>
<td>18.4</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>1234</td>
<td>1234</td>
</tr>
</tbody>
</table>
No Observer condition is slight, there is a clear trend towards shorter average durations as the Observer’s alignment shifts towards Player A. We also note that in every condition, the intended stopping times are always longer than the actual waiting times (which is expected since the actual waiting time is the minimum of the two players’ intended waiting times when both times are implemented). In addition, we find that Player A’s intended stopping times are longer than Player B’s. These patterns provide some support for the expectations stated in Hypothesis 2.

The regression analysis in Table 2 provides rigorous statistical support for these findings. We regress the actual and intended stopping times on an indicator variable for the Observer condition (to measure the effect of the presence of any observer compared to no observer) and separate indicators for whether the observer’s ideal point is 25 or 0 (to measure the effect of alignment between the Observer and Player A relative to the ideal point of 50). The direction of the effect of the Observer condition on actual stopping times is consistent with Hypothesis 2, and although none of the coefficients in the first model are individually significant, they are jointly significant ($p = 0.03$). Furthermore, the linear combinations of the coefficients imply that the stopping times in the Observer 25 and Observer 0 conditions are 1.87 seconds shorter ($p = 0.06$) and 2.67 seconds shorter ($p < 0.01$), respectively, than in the baseline No Observer condition. This latter finding lends support for Hypothesis 3. Together, these results indicate that stopping times decrease once an Observer is introduced and that the effect is more pronounced when there is greater alignment between the Observer and Player A’s interests.

When we turn from analyzing average waiting times to the shapes of the distributions in the observer conditions, as shown in Figure 2, we do see differences between the conditions in terms of the proportion of early and late concessions. The number of players conceding very early in the contest appears to be higher under all observer treatments than under the No Observer treatment. Looking at the right-most distribution, early concessions more
Table 2: Effects of Treatment on Contestant Choice of Stopping Times

<table>
<thead>
<tr>
<th></th>
<th>(1) Actual</th>
<th>(2) Intended: Player A</th>
<th>(3) Intended: Player B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer Present</td>
<td>-1.15</td>
<td>-1.55*</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.63)</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Observer 25</td>
<td>-0.72</td>
<td>0.40</td>
<td>-1.31+</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(0.77)</td>
<td>(0.77)</td>
</tr>
<tr>
<td>Observer 0</td>
<td>-1.51</td>
<td>-1.13</td>
<td>-1.43+</td>
</tr>
<tr>
<td></td>
<td>(1.19)</td>
<td>(0.77)</td>
<td>(0.77)</td>
</tr>
<tr>
<td>Constant</td>
<td>15.40**</td>
<td>19.42**</td>
<td>17.19**</td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td>(0.31)</td>
<td>(0.31)</td>
</tr>
</tbody>
</table>

N 1235 1326 1326
R² 0.01 0.02 0.01

Standard errors in parentheses
No Observer treatment omitted as reference category.
Models (2) and (3) use subject fixed effects.
+ p < 0.10 * p < 0.05, ** p < 0.01

than double when the alignment between the observer and Player A is greatest. Subjects also shift their concession times from the middle of the game to either extreme when the Observer’s ideal point is 0, as the proportion of rounds with immediate concessions ($t = 0$) and last-possible moment concessions ($t = 30$) are highest in this condition than in any other. Kolmogorov-Smirnov equality of distribution tests show that the distribution under the No Observer treatment is significantly different from each of the observer treatments ($p < 0.01$ for each treatment).

The interpretation of these visual patterns is supported by additional regression analysis. In Table 3, we present estimates for a series of linear probability models in which we characterize stopping times as either early or late. The model in the left-most column shows that in the baseline No Observer condition, 11% of contests end immediately. When an observer is added, the chance of the game ending immediately increases by 6 percentage points. More important, this probability increases by another 7 percentage points when the
Figure 2: Distribution of Actual Stopping Times

![Distribution of Actual Stopping Times](image)
alignment between the observer and Player A is greatest. Indeed, the probability of the game ending immediately is more than twice as likely when the observer’s ideal point is 0 than when there is no observer. This finding provides additional qualified support for Hypothesis 2 and Hypothesis 3.

Table 3: Effects of Treatment on Likelihood of An Extreme Stopping Time

<table>
<thead>
<tr>
<th></th>
<th>Immediately</th>
<th>First 5 Seconds</th>
<th>Last 5 Seconds</th>
<th>Last Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer Present</td>
<td>0.06*</td>
<td>0.06</td>
<td>-0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Observer 25</td>
<td>-0.00</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Observer 0</td>
<td>0.07*</td>
<td>0.02</td>
<td>-0.05</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.11**</td>
<td>0.31**</td>
<td>0.36**</td>
<td>0.90**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

N 1234 1234 1234 1234
$R^2$ 0.017 0.006 0.003 0.002

Standard errors in parentheses
† $p<0.10$, ∗ $p<0.05$, ∗∗ $p<0.01$

We also find that the outcome of the Staring Contest (in terms of whether A wins, B wins, or there is disagreement) depends on the presence and preferences of the observer. Figure 3 shows the distribution of these outcomes by observer condition. Two results are noteworthy. First, we see that breakdown is less frequent in rounds where there is an observer than in rounds without one. This provides further support for Hypothesis 2. Second, the more interesting and clearly visible result is the shift in outcomes as a function of the observer’s ideal point. Although Player B wins more often when the observer’s ideal point is equidistant from both players’ ideal points (Obs 50), as the observer’s ideal point becomes closer to Player A’s ideal point, Player A wins the game more often, consistent with Hypothesis 3. Indeed, we see that Player A has the greatest advantage over Player B in Obs
0 than in any other conditions: Player A wins 54.2% of contests compared to Player B, who wins 40.3% of them.

Figure 3: Distribution of Outcomes by Treatment

The extent to which the Observer’s presence and alignment with Player A affects the outcome of the Contest is supported by the linear probability model estimates in Table 4. The presence of the observer implies a statistically significant decrease of 4 percentage points in the probability of breakdown (estimates in the first column). When the alignment between the Observer and Player A is greatest (Obs 0), the likelihood that Player B wins is 10 percentage points lower than when there is no alignment (Obs 50), and the likelihood Player A wins is 9 percentage points higher. Thus, we find clear evidence the alignment between the observer and Player A has a significant effect on who wins the Staring Contest, consistent with the expectation in Hypothesis 5.
Table 4: **Effect of Observer Presence, and Alignment with Player A**

<table>
<thead>
<tr>
<th></th>
<th>Breakdown</th>
<th>B Wins</th>
<th>A Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer Present</td>
<td>-0.04*</td>
<td>0.08*</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Observer 25</td>
<td>0.02</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Observer 0</td>
<td>0.01</td>
<td>-0.10*</td>
<td>0.09+</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.05)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.08**</td>
<td>0.42**</td>
<td>0.49**</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

N 1326  1326  1326  
R² 0.004  0.004  0.003

Standard errors in parentheses
Observer Ideal 50 omitted as reference category.
+ p < 0.10, * p < 0.05, **p < 0.01

Our analysis thus indicates that there are two primary effects of introducing an Observer into the Starting Contest. First, the duration of the Staring Contest decreases and is less likely to break down. This is consistent with the possibility that observers may use punishment to encourage contestants to resolve standoffs more quickly. Second, the degree of alignment between the Observer and Player A markedly influences who wins and who loses the standoff. When the Observer is aligned with Player A, Player A clearly wins the contest more often, which is consistent with our hypothesis that observers can try to sway the outcome by punishing Player B more heavily than Player A. We take a closer look at how observers punish the contestants in the next section.

### 4.2 Punishment

Table 5 displays the average number of points the observer deducts from each contestant by outcome and treatment condition. The pattern of punishment behavior appears to lend support for our hypotheses regarding instrumental punishment. First, observers punish the
winning player more than they punish the losing player. When the observer’s ideal point is 50 so that there is no preference between A winning and B winning, the observer deducts 23.7-29.9 points from the winner compared to 8.9-10 points from the loser. In this case, the deductions to Player A and Player B are not significantly different conditional on winning \((p = 0.20)\) or losing \((p = 0.72)\). These results therefore provide qualitative support for Hypothesis 4.

Second, when there is alignment between the observer and Player A, we find that the observer punishes Player B for winning much more heavily than Player A. The average deductions to Player B for winning are 32.8 (Obs 25) and 38.7 (Obs 0), while the deductions to A for winning are 22.4 (Obs 25) and 19.9 (Obs 0). For both Obs 25 and Obs 0, the punishment received by Players A and B for winning are significantly different \((p = 0.03\) and \(p < 0.01)\), lending support for Hypothesis 5.

While the tendency to punish the winner is qualitatively consistent with our theoretical expectations, there are several aspects of punishment behavior that we did not anticipate and that appear to be less than optimal. First, the magnitude of punishment is weaker than we expected. Recall that punishment of at least 80 points to the winner, if anticipated by the Staring Contest players, could induce a change in behavior and therefore in the outcome of the contest. The level of point deductions tend to be far less, on average, than 80 points.

Table 5: Average Deductions by Contest, Outcomes, and Treatment

<table>
<thead>
<tr>
<th>Contest Outcome</th>
<th>Observer Ideal Point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>A Wins Punishment to A</td>
<td>19.9</td>
</tr>
<tr>
<td>A Wins Punishment to B</td>
<td>6.4</td>
</tr>
<tr>
<td>B Wins Punishment to A</td>
<td>18.8</td>
</tr>
<tr>
<td>B Wins Punishment to B</td>
<td>38.7</td>
</tr>
<tr>
<td>Breakdown Punishment to A</td>
<td>38.3</td>
</tr>
<tr>
<td>Breakdown Punishment to B</td>
<td>43.3</td>
</tr>
</tbody>
</table>
Second, observers punish contest *losers* in addition to contest winners. Such punishment does not alter the direction of the contestants’ incentives—it simply makes losing worse than it already is. Third, while observers punish winners more than losers, Player A’s punishment seems to be *insensitive* to winning or losing in Obs 25 and Obs 0; in these conditions, only Player B’s punishment is sensitive to winning or losing. Finally, observers tend to punish *both* contestants heavily in the case of breakdown. When neither player concedes in the Star-ing Contest, the observer deducts between 24.4 and 41.1 points from Player A and between 24.4 and 43.3 points from Player B. These deductions are not significantly different for the two players given the observer’s ideal point and, moreover, do not appear to depend on the alignment of interest between the observer and Player A. Taken together, these unanticipated patterns of behavior may reflect features of blame that have an instinctive or emotional basis rather than a fully instrumental one.

Regression analysis in Table 6 provides another way of looking at observer behavior by examining punishment to the winner and loser separately. In this analysis, we regress the amount of points deducted on an indicator for whether Player B wins and a control for the actual stopping time. Several patterns clearly arise. First, the single largest estimate in the table is the constant for punishment of the Winner when the Observer’s ideal point is 50. This indicates that neutral Observers choose to blame the winner heavily, regardless of which player won. This is direct evidence supporting Hypothesis 4. That said, introducing any alignment between Player A and the Observer shifts punishment towards Player B. In the Obs 50 condition, punishment to the winner and loser does not depend on which player wins, but it does in the Obs 25 and Obs 0 conditions. When Player B wins in the Obs 25 and Obs 0 conditions, Player B is punished more heavily as the winner than Player A is as the winner. Interestingly, Player A is also punished more heavily for losing, though not as much as Player B is punished for winning. This analysis provides further support for the interpretation that observers use punishment not only instrumentally, but also as a
Table 6: Effects of Observer Ideal Points on Punishment to Winner and Loser

<table>
<thead>
<tr>
<th>Points Deducted from Winner</th>
<th>Observer Ideal 0</th>
<th>Observer Ideal 25</th>
<th>Observer Ideal 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Wins</td>
<td>20.64**</td>
<td>12.97**</td>
<td>-3.74</td>
</tr>
<tr>
<td></td>
<td>(4.88)</td>
<td>(4.63)</td>
<td>(4.73)</td>
</tr>
<tr>
<td>Actual Stopping Time</td>
<td>-0.14</td>
<td>0.14</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>Constant</td>
<td>19.89**</td>
<td>17.69**</td>
<td>28.15**</td>
</tr>
<tr>
<td></td>
<td>(4.06)</td>
<td>(4.07)</td>
<td>(4.43)</td>
</tr>
<tr>
<td>N</td>
<td>216</td>
<td>216</td>
<td>216</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.08</td>
<td>0.04</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Points Deducted from Loser</th>
<th>Observer Ideal 0</th>
<th>Observer Ideal 25</th>
<th>Observer Ideal 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Wins</td>
<td>13.03**</td>
<td>15.46**</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td>(3.14)</td>
<td>(3.24)</td>
<td>(2.81)</td>
</tr>
<tr>
<td>Actual Stopping Time</td>
<td>-0.22$^+$</td>
<td>-0.13</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.13)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Constant</td>
<td>8.83**</td>
<td>7.41**</td>
<td>8.01**</td>
</tr>
<tr>
<td></td>
<td>(2.62)</td>
<td>(2.85)</td>
<td>(2.63)</td>
</tr>
<tr>
<td>N</td>
<td>216</td>
<td>216</td>
<td>216</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.09</td>
<td>0.10</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Standard errors in parentheses.
$^+ p < 0.10, * p < 0.05, ** p < 0.01$
simple way to express their displeasure with the outcome, consistent with the notion that punishment is a behavioral form of blame.

One might wonder whether punishing the winner can be entirely explained by inequity aversion, rather than instrumental punishment or behavioral blame (Hatfield et al., 1978; Fehr and Schmidt, 1999; Goeree and Holt, 2000; Engelmann and Strobel, 2004). Such an argument implies the observer deducts points in order to equalize the payoffs of the winner and loser of the Staring Contest. Although this might seem plausible at first glance, there are several reasons why we do not believe inequity aversion can completely explain the observer’s behavior. First, punishing both players in the case of disagreement increases, rather than decreases, inequality between the contestants and the observer. We see no particular reason why an inequity averse subject would care only about inequality between the contestants but not inequality between the contestants and the observer. Second, punishing the loser is inconsistent with inequity aversion, as doing so only serves to increase inequality for any given level of punishment to the winner. Third, inequity aversion cannot explain why punishment is asymmetric in the Obs 25 and Obs 0 conditions. That is, it does not explain why Player B’s punishment in these conditions is conditional on winning while Player A’s punishment does not depend on winning or losing. Overall, blame seems to reflect both instrumental motivations as well as emotional or instinctive reactions that cannot be fully accounted for by concerns for distributional equality.

### 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 rare, most punishment comports with our expectation that observers will 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 $\theta_0$). 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 these 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 7 presents these results. Categories are separated by the outcome of the game to show that strategies appear dependent on the outcome of the Contest.

We first note that subjects do not shy away from punishing the contestants. Positive punishment is used in over half of all periods, and 83% of subjects use some punishment in at least one round of the experiment. When punishment is non-zero, punishing the winner
Table 7: Distribution of Play and Subjects by Category

<table>
<thead>
<tr>
<th>Punishment Strategy</th>
<th>% of Rounds</th>
<th>% by Outcome</th>
<th>% of Subjects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Breakdown</td>
<td>A Wins</td>
<td>B Wins</td>
</tr>
<tr>
<td>Strongly Rational</td>
<td>11%</td>
<td>*</td>
<td>6%</td>
<td>17%</td>
</tr>
<tr>
<td>Weakly Rational</td>
<td>25%</td>
<td>*</td>
<td>13%</td>
<td>37%</td>
</tr>
<tr>
<td>Punish the Winner</td>
<td>34%</td>
<td>*</td>
<td>35%</td>
<td>37%</td>
</tr>
<tr>
<td>Equal</td>
<td>14%</td>
<td>57%</td>
<td>9%</td>
<td>14%</td>
</tr>
<tr>
<td>None</td>
<td>42%</td>
<td>20%</td>
<td>50%</td>
<td>37%</td>
</tr>
</tbody>
</table>

*Under Breakdown, there is no winner to punish, and thus no results for these categories.

appears to be the prevailing strategy.\(^{16}\) Although a small percentage of observations fit the requirements of strongly rational play (11%), we do find that a quarter of observations can be classified as weakly rational (25%) and that even more rounds are consistent with punishing the winner (36%). When we condition on the use of non-zero punishment, 20% of such observations can be classified as strongly rational, 44% count as weakly rational, and almost two-thirds (64%) are consistent with punishing the winner. We observe very few rounds with equal punishment (12%), while the observer uses no punishment whatsoever in slightly less than half of the rounds (44%). The subject-level classifications are generally consistent with the observation-level findings except that few subjects consistently use strongly rational punishment.

Observers exhibit a tendency to use rational allocations more often when Player B wins the Contest. Both the strong and weak versions of rational play occur three times as often when Player B wins (17% and 37%) than when Player A wins (6% and 13%). This is consistent with the asymmetry in punishment in Obs 25 and Obs 0 discussed above. We also note that any punishment is used less frequently when Player A wins (50%) than when

\(^{16}\)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 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[.]”

29
Player B wins (63%). These results highlight the importance of observer-player alignment and the allocation of blame.

5 Discussion

How can blame be used as a mechanism for citizens to influence their representatives? In our experiment, we investigated how observers apportion blame and how the anticipation of blame affects political outcomes, finding that observers used blame as punishment for delaying the resolution of political standoffs. Contest winners are punished more often than losers, and enemies are punished more heavily than friends. However, friends are also punished when they lose. Blame is neither indiscriminately placed nor used solely to scold the opposition. These results suggest that political representatives may not necessarily be immune to being blamed by voters who share their partisanship or preferences—recalcitrance will still be punished. While these findings generally comport with instrumental rationality, observers’ punishment behavior only weakly affects the incentives for contestants.

We also find that blame is consequential. Observers appear to be able to shape the contestants’ behavior in ways that increase the chances of their preferred outcome. The presence of the observer reduces the duration of the contests, and the asymmetric punishment levied against opponents tilts outcomes in favor aligned contestants. Although blame may reflect both an instrumental motive as well as a simple emotional or behavioral response, contestants are nevertheless sensitive to weak forms of blame.

Extensions to our work would move beyond our focus on how voter blame impacts negotiating behavior during standoffs that are already underway. The single-shot game highlighted how an audience to the standoff decides to allocate punishment at a single point in time. Of course, political negotiations are 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.\textsuperscript{17} Prompt resolution of subsequent debt limit battles convinced another ratings firm to return the U.S. credit rating to full AAA status.\textsuperscript{18} 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.\textsuperscript{19} Speaker John Boehner over-estimated 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.\textsuperscript{20} Following these episodes, Gingrich and Boehner seemed to learn a lesson, as later battles ended in favor of their opponents.\textsuperscript{21}

\textbf{References}


\textsuperscript{18}Jim Puzzanghera, “Fitch takes U.S. credit rating off downgrade watch after debt deal”, \textit{Los Angeles Times}, 21 March, 2014.
\textsuperscript{20}Paul Steinhauser, “CNN Poll: GOP would bear the brunt of shutdown blame”, \textit{CNN Politics}, 30 September 2013.


33


