

Running Head: DETERRENCE AND COERCIVE CAPACITY

Does Aggression Deter or Invite Reciprocal Behavior? Considering Coercive Capacity

Stephen Benard

Department of Sociology
Indiana University

Mark T. Berg

Department of Sociology & Public Policy Center
University of Iowa

Trenton D. Mize

Department of Sociology & Cluster in Advanced Social Science Methodologies
Purdue University

Corresponding Author:

Stephen Benard, Department of Sociology, Indiana University, 744 Ballantine Hall,
Bloomington, IN 47405.

Email: sbenard@indiana.edu

ABSTRACT

How do people respond to aggression? Theory differs on whether aggressive behavior deters antagonists or provokes retaliation, and the empirical evidence is mixed. We bridge contradictory findings in the literature by identifying a previously unexamined moderating variable: the extent to which individuals can increase their coercive capability (which we call “escalating”). We argue that when escalating is costly, aggression deters potential antagonists. In contrast, when escalating is less costly, behaving aggressively fails to deter aggressive partners. We test these predictions in two behavioral experiments that manipulate the cost of escalating and whether interaction partners are aggressive or deferential. We find support for deterrence predictions when escalating is either high or low cost, but not when it is medium cost. Taken together, we provide evidence that the cost of escalation plays a key role in decisions about aggression.

When state power is absent, weak, or mistrusted, as in rural “honor cultures”, disenfranchised urban neighborhoods, or international relations, individuals cannot rely on authorities to enforce the law. Instead, they devise other strategies to safely navigate a social landscape marked by the potential for exploitation and loss. One of the most common strategies is to display aggressive and threatening behavior. By behaving aggressively, individuals signal that they possess the coercive capability – such as physical toughness or access to weapons – to prevail if an interaction reaches conflict. In theory, this deters exploitation from potential adversaries, who opt for less threatening targets.

Despite much cross-disciplinary research, it is unclear whether this strategy actually deters aggression. Past work suggests two possibilities. First, by signaling coercive capability, aggression may serve as a credible threat that deters aggression from others. Conversely, by eliciting reciprocity, aggression may provoke revenge rather than deference. Empirical studies similarly diverge. A number of laboratory experiments find that the capacity for aggression (Lawler et al. 1988), the use of coercive power (Molm 1997), and aggressive behavior (Benard 2013; Benard 2015) *deter* aggression from others. In contrast, a large body of observational research supports *reciprocity* arguments as aggressive individuals suffer relatively higher rates of victimization (e.g., Berg and Loeber 2015; Sampson and Lauritsen 1994). We suggest these contradictory findings can in part be explained by a previously unexamined moderating factor: the extent to which individuals can increase their coercive capability. While coercive capability is often held constant in laboratory experiments, in many real world settings individuals can increase their coercive capability (e.g., by arming oneself) – what we refer to as escalation. Many accounts of homicide, for example, describe verbal altercations that end violently when one disputant escalates the situation by retrieving a weapon. Increases in coercive capability alter the

power equation in hostile encounters, enabling weaker opponents to match or exceed the coercive capability of their adversaries. For example, Gould (2003:75) describes a dispute in 1850s Corsica in which a smaller man escalated a dispute with a larger man by threatening the latter with first a rock and then a knife, and was disarmed in both cases. He then left the scene, borrowed a rifle, and returned to fatally shoot the larger man.

We argue that when increasing coercive capacity is costly or impossible for all individuals in a particular situation – as in most experimental studies and some real world settings—behaving aggressively deters aggression from others. In contrast, when it is less costly for individuals to increase coercive capacity – as in many real-world settings studied by criminologists – behaving aggressively fails to deter aggressive partners, because those partners can increase their coercive capacity to match that of their adversary.¹ By identifying the cost of increasing coercive capability as a moderating variable that explains when aggression produces deference or retaliation, we seek to explain a portion of the divergence in past theory and research, and contribute to the development of a more general framework for understanding aggressive behavior. In the present studies, we focus on settings in which escalation allows an individual to match the coercive capacity of their partner. We would expect escalation to do more to facilitate retaliation when it allows actors to exceed their partner’s coercive capacity, and less if it does not allow actors to match their partner’s capacities.

Broadly, we examine the following questions: (1) when an individual is the target of aggression, do they respond to the antagonist by increasing their coercive capabilities? (2) When individuals can more easily increase their coercive capabilities, will their adversary’s aggressive behavior deter or instead invite revenge? Theoretically, answering these questions requires

understanding how actors weigh factors such as their relative coercive capability, concerns about deterrence, moralistic beliefs about justice, and their own emotions.

THEORY

Aggression—both violent and non-violent—is often goal-oriented: it serves to rectify perceived injustices, but also to gain compliance in predatory exchanges. Although aggression is often costly—in some instances resulting in banishment, serious injury, or worse—it commonly serves as a route to contesting power, status, and resources in diverse settings, including adolescent peer groups (Faris 2012), organizations (Loch, Yaziji, and Langen 2000), markets (Clark and Montgomery 1998), and international relations (Bueno de Mesquita et. al 1997).

An important factor shaping success in such contests is an actor's *coercive capability*, or their capacity to prevail in conflict (Crowley 2000; Johnstone and Bshary 2004; Lawler, Ford, and Blegen 1988). Coercive capability takes many forms, such as physical toughness and access to weapons in violent disputes (Anderson 2000), or the ability to deploy advertising, lawsuits, or price wars in contests between firms (Clark and Montgomery 1998; Podolny and Scott Morton 1999). Generally, those with greater coercive capability prevail in competitive encounters.

Actors can often observe their own coercive ability, but not the coercive capability of others. For example, armed robbers often do not know if a potential victim is carrying a gun (Jacobs 2000), and firms often cannot directly observe their rivals' economic competitiveness (Clark and Montgomery 1998:65). This creates ambiguity, as actors are unsure of their chances of winning a contest, and thus whether it pays to engage in conflict. This ambiguity can motivate aggression, either because actors attempt to intimidate their rivals (Benard 2013; Schelling 1960), or because they evaluate their rivals' coercive capability ability by provoking them to see if they will retaliate (Gould 2003). Coercive capability plays an important role in both in formal

models of aggression (Crowley 2000; Johnstone and Bshary 2004) and implicitly in less formal discussions of the importance of displaying toughness or other coercive mechanisms (Nisbett and Cohen 1996; Schelling 1960).

Our goals are to understand how the cost of adjusting coercive capacity and partner aggression affects actors' willingness to (1) increase their coercive capability and (2) to aggress towards or challenge their opponents. Because our interest lies in testing the theoretical implications of these differences, we employ a laboratory setting in which increasing coercive capability is high versus low cost (in study 2, we lower the costs further to examine a broader range of the parameter space). We use cost as a proxy for the myriad factors that affect the ease of adjusting coercive capability. For concision, we use the term "escalate" interchangeably with the more cumbersome phrase "paying to increase coercive capacity." We expect that when the cost of escalating is higher, actors will be less likely to do so. Thus:

***Hypothesis 1:** Individuals will be less likely to escalate when the cost of doing so is higher.*

Similarly, we expect that when the cost escalating is higher, individuals will be less likely to aggress against opponents overall. This is because they will be less likely to possess high coercive capability themselves:

***Hypothesis 2:** Individuals will be less likely to aggress when the cost of escalating is higher.*

We examine how a partner's aggressive or deferential behavior shapes an individual's decision to escalate. When facing an aggressive partner, individuals should be more likely to assume that their partner has escalated, and be more willing to escalate in order to protect themselves:

Hypothesis 3: *When an interaction partner is aggressive (versus deferential) towards an individual, the individual will be more likely to escalate.*

While the hypothesized direct effects of the cost of escalation and partner behavior are straightforward, we expect that cost will play a more theoretically interesting role as a moderating variable affecting individuals' responses to aggression from partners in competitive exchanges. To develop hypotheses regarding these moderating effects, we draw on interdisciplinary theory and research from both the deterrence and reciprocity perspectives.

Perspectives on Deterrence

One line of research posits that individuals are deferential in response to aggression from others because that aggression is taken as a sign of the interaction partner's coercive capacity and willingness to harm the individual. For example, Schelling (1960) argued that credible threats signal to one's bargaining partner that the partner faces great costs if they do not defer. One of the clearest ways to establish the credibility of a threat is to have acted on similar threats in the past (Frank 1988; Gould 2003; Schelling 1960). A number of experimental studies support deterrence predictions, finding that coercive capability (Lawler, Ford, and Blegen 1988), strong, consistent, coercive punishments (Molm 1997), and both direct aggression from a partner and a reputation for past aggression (Benard 2013; 2015) elicit cooperation and deference from interaction partners. More broadly, ethnographic work and case studies on "honor cultures" and the "code of the street" finds that individuals often testify to the deterrent value of aggression (Anderson 2000; Beckerman et. al 2009).

Perspectives on Reciprocity

Reciprocity is a fundamental dimension of human behavior (Blau 1964; Molm 2010), and these reciprocal tendencies extend to aggression. In lab studies, individuals reciprocate electric shocks

(Helm, Bonoma, Tedeschi 1972) and spend money to punish uncooperative partners (Fehr & Gächter 2002; Yamagishi 1986). Outside of the lab, reciprocity concerns underlie a range of behaviors, from gang related homicide (Papachristos 2009), violence in illegal drug markets (Topalli, Wright and Fornango 2002), to more pedestrian forms of workplace retaliation (Aquino, Tripp, and Bies 1999). Reciprocity may stem from several sources. Competitive encounters generate strong emotions that overwhelm inhibitions (e.g. Domsio 1994; Katz 1988; McDermott 2004), and may render individuals more willing to harm others (Ariely and Lowenstein 2006). Concerns about justice or a sense of moral outrage can also lead actors to reciprocate aggression, even when the costs outweigh the material benefits (Jacobs and Wright 2006:30). Aside from punishment and justice motives, reciprocity also is motivated by efforts at assertive self-presentation (Tedeschi and Felson 1994).

When Does Aggression Deter versus Invite Reciprocal Aggression?

To integrate these two lines of research, we propose that individuals feel a desire to retaliate when they are the targets of aggression, but hesitate to act on these desires for fear of incurring additional costs, such as punishment, guilt, stigma, damaged social relationships, or further harm if the revenge attempt is unsuccessful. When individuals can easily enhance their coercive capability, this can temper concerns about suffering harm, because it improves the odds that their retaliatory efforts will succeed. For example, firearms enable physically weaker individuals to incapacitate more dangerous opponents whom they normally would not challenge for fear of retribution (Felson, Berg, and Rogers 2014). When there is an initial power imbalance among antagonists, the weaker opponent can match or exceed the capabilities of their adversary, leading to confrontation (Cook and Ludwig 2006). We expect that the cost of increasing coercive capability moderates the effect of aggressive behavior on the likelihood of reciprocal aggression.

Aggressive partners are more likely to be perceived as having chosen to escalate, and when the cost of escalating is higher, individuals are less likely to reciprocate. Specifically, we propose that:

***Hypothesis 4:** When the cost of escalating is sufficiently high, aggressive behavior from partners will have a deterrent effect, such that individuals are less likely to aggress toward partners who have been aggressive towards them previously (compared to those who have not).*

***Hypothesis 5:** When the cost of escalating is sufficiently low, the deterrent effect of aggression will be reversed, such that targets will be more likely to reciprocate aggression.*

Hypotheses 4 and 5 predict that the deterrent effect of aggression will be weaker in settings when the cost of increasing coercive capability is lower.

Design

To test our hypotheses, we need to compare how individuals respond to aggressive or deferential partners, while systematically varying the cost of increasing coercive capacity for all individuals. To this end, we conduct two behavioral experiments in which individuals interact in a lab setting commonly used to study aggressive behavior. Individuals are randomly paired with different partners over a series of rounds, and on each round, choose whether to escalate, and whether to behave aggressively or deferentially towards their partner. To determine the causal influence of our factors of interest, we experimentally manipulate the cost of escalating (medium versus high cost in Study 1, low versus medium cost in Study 2) as well as the aggressive behavior of participants' interaction partners.

STUDY 1

Participants

The study included 120 paid undergraduate volunteers (42 men and 78 women) at a large public university in the Midwestern US.² Participants were recruited via advertisements offering payment for participation in an experimental study. A total of 11 participants (9.2%) were removed from the analysis due to suspicion about some aspect of the study procedure (e.g., correctly guessing that their partners were simulated). Suspicion was assessed in the post-study interview, using a funneled debriefing procedure. Due to a computer error, we are missing final round data for 29 cases (<1% of the total observations). Experimental sessions generally included 6 participants, in sessions with fewer participants, confederates filled in to create the appearance of a full session. Confederates sat at a nearby computer to simulate the presence of an additional participant; because partners were simulated, the confederates did not interact with the participants, and so could not affect their responses.

Procedure

When participants arrived at the lab, a research assistant showed them to a private cubicle equipped with a networked computer. Participants were led into the room one at a time, and curtains were drawn over the cubicle entrance to prevent participants from seeing one another. While participants could not see one another, they could hear other participants typing, which should increase realism and decrease the perception that partners are simulated. While participants believed they were interacting with other participants in the room via the computer software, their partners' responses were in fact simulated by the computer program. Simulating partners is necessary for testing the hypotheses, as it allows us to systematically vary the behavior of participants' interaction partners. After receiving both written and verbal instructions

describing the study procedures, all participants took a quiz to ensure they understood the study instructions, and were encouraged to ask questions if any aspects of the study were unclear.

The study took place over 35 rounds, plus 5 practice rounds. On each round, participants interacted in a generalized hawk-dove game, designed to model the incentive structure of competitive environments in which participants choose between aggressive and deferential options (described below). Participants earned points for their decisions, and believed that accumulating more points would translate to greater earnings. This provided participants with an incentive to take their decisions seriously. In fact, all participants were paid the maximum (\$15), because simulating their partners meant that their point totals were in part determined by the experimenter. Following the behavioral portion of the study, participants responded to a survey that included demographic questions as well as some additional measures.³ Following the study, participants were debriefed, paid, and thanked for their participation in a private room to protect their confidentiality.

The Generalized Hawk-Dove Game

To test the hypotheses, we required a setting that would allow participants to (1) choose between deferential and aggressive behavior, with some risk of exploitation, (2) decide whether to escalate, (3) interact without a central authority regulating aggression, (4) observe their own competitive ability, but (5) not observe the competitive ability of their partners. To meet these requirements, we modified the generalized hawk-dove game (GHD), a longstanding and widely used model of conflict and aggression, to allow participants to escalate (Crowley 2000; see also Halevy et al 2006; Snyder 1971; Bowles and Gintis 1993). The GHD is depicted in the top panel of Table 1.

[Table 1 about Here]

On each round of the study, participants are paired with a simulated partner, ostensibly from the group in the lab, for a one-shot interaction in the GHD. The GHD models a competition for a pool of resources with a value V . Participants can choose to challenge and claim all of the resources, or defer, and allow their partner to take the lead in dividing the resources. If one player challenges and the other defers, the unilateral challenger receives the all of the resource (earning V), while their deferential partner earns 0 (shown in the lower left and upper right cells of Table 1). If the interaction partners mutually defer, they split the resource, minus a cost T (representing costs of negotiating, see the lower right cell in Table 1). If the participants mutually challenge (both attempt to claim the resource), the situation escalates to conflict. In this case, one participant wins with probability p . The winner receives the entire resource, minus a cost of winning W . The individual losing the challenge receives no resources, and additionally pays a cost of losing L . The L parameter represents costs such as injuries, court costs, or other losses. The inclusion of competition, with the risk of losing more than one stands to gain distinguishes the GHD from purely distributional or other zero sum settings.

In this study, we set $V = 50$, $W = 0$, $L = 100$, and $T = 10$. This means that winners and unilateral challengers earn 50 points, losing a competition costs 100, mutual deference earns both partners 15 points, and unilateral deference earn 0 points. The bottom panel of Table 1 shows the payoff table that was shown to participants during Study 1. All participants began the study with an initial endowment of 500 points to avoid negative earnings.

The game was implemented via a computer program (programmed in zTree; Fischbacher 2007). Participants' screens displayed several pieces of information at the beginning of the round. They were shown a table illustrating the game, including payoffs for each combination of choices. They were also reminded of whether their partner challenged them on their last

interaction (e.g. “In your last meeting, participant C: challenged”), and a list indicating whether their partner had paid to increase their coercive capability on each of the past 5 rounds.

Participants viewed this information while making two decisions. First, they decided whether to escalate. In the GHD game, coercive capability determines an individual’s probability of winning if both players challenge. For simplicity, participants could be high or low in coercive capability. All participants began each round low in coercive capability, and could choose to pay a certain amount (which varied by condition, described below in the independent variables section) to have high coercive capability for that round. In the study, the instructions referred to coercive capability as the participants’ “power rating,” because of potential social desirability issues with the term “coercive.” Participants were told the actual cost of escalating, without the use of any relative labels (e.g. “high” or “low”) that might have influenced their perceptions of these costs. The text read: “Your current power rating is low. You can choose to pay X points to raise your power rating and increase your chance of winning if both players challenge. Would you like to pay X points to raise your power rating?” The instructions clearly stated that all participants in the setting had the opportunity to escalate on each round. Second, at the same time participants decided whether to escalate, they also decided whether to challenge their partner, or defer.

Independent Variables

There are two key independent variables in the study: the cost of escalating and the aggressiveness of each participant’s simulated partner. In addition, we vary each partner’s prior reputation for increasing coercive capability. This allows us to keep our design similar to several prior studies and replicate prior findings regarding the effects of reputations for aggressive behavior (e.g. Benard 2013, 2015). We also include a control for time.

Cost of escalation/ increasing coercive capability: Participants were randomly assigned to one of two between-subjects conditions that varied on the amount participants needed to invest in order to escalate from low to high for one round. One challenge in operationally defining this variable lies in determining what participants will experience as a “high” or “low” cost. In the high cost condition, escalating cost 75 points. We chose this value because it is a 50% increase over the maximum amount participants stood to gain from each interaction. High-cost escalation guarantees that the participant will lose points, even if they win the competition; we expect they will experience it as unambiguously high cost.

In the other condition, participants needed to invest 25 points to escalate for one round. We chose this value because it is 50% less than the maximum value participants stand to earn from competitive interactions. Participants can escalate and still profit if they win, though less than if they win without escalating. While our initial goal was to formulate a low-cost condition, we later (Study 2) introduce a still lower cost condition as an additional comparison. Thus, for concision, we refer to the lower-cost condition in Study 1 as the “medium cost” condition, and the condition introduced in Study 2 as the “low cost” condition.

Participants were told that they had a 50% chance of winning a mutual challenge if both partners had the same level of coercive capability, and a 75% or 25% chance of winning if they were higher or lower in coercive capability, respectively. We implemented this in the study by assigning participants who chose to escalate a 62.5% chance of winning, and those who did not a 37.5% chance, which represents their chances of winning assuming half of their partners escalate.⁴

Partner aggression: We systematically varied the aggressiveness of participants’ simulated partners across the 35 rounds. Partners were labeled (with letters during the main part of the

study and numbers during the practice rounds) so participants could recall whether their partner challenged them on their last interaction (referred to here, but not in the experiment instructions, as “aggressive partners”) or deferred to them on their last interaction (“not aggressive”). In addition, rather than relying on participants’ memories, we reminded them each round whether their partner had behaved aggressively or non-aggressively on their previous meeting.

Partner reputation: Participants were also able to view their partner’s overall reputation for increasing their coercive capability, indicating how many times the partner had chosen to increase their coercive capability in the last five rounds in interactions with other opponents.⁵ To make the behavior of partners appear less random and more realistic, specific simulated partners were presented as increasing their coercive capability often (two partners consistently escalated four to five times in the last five rounds), rarely (two partners escalated 0-1 times in the last five rounds), or a mix (one partner escalated 2-3 times in the last five rounds). We crossed these partner types with the partner aggression variable, so that of the frequent and infrequent increasers, one was aggressive (always challenge) and one was non-aggressive (always defer). The final partner mixed challenging and deferring, consistent with their record of increasing coercive ability a portion of the time. For concision, the effects of partner reputation are discussed in Online Supplementary Appendix S1.

Time Block: Lastly, we controlled for the effect of time block of the experiment. All participants participated in 35 non-practice rounds, being randomly paired with one of the five partners described above in blocks of five rounds (therefore each participant was paired with each partner once every five rounds). This produced seven total time blocks. We exclude the first time block from the analysis, as participants lacked full information on their partners’ past behavior in this time block. Time block was specified as a set of dummy variables to account for possible

nonlinear effects of time (with block two, the first block used in the analysis, as the reference category). As a sensitivity check, we fit alternative models (1) including the first time block in the analysis and (2) specifying time as a continuous variable rather than a series of dummy variables. Our substantive conclusions are unchanged using these alternative modeling strategies. We omit the estimates for the time block variables in the presented tables to preserve space as we have no hypotheses regarding time; readers interested in the effect of time can find full tables including the coefficients for the time block variables as well as graphical depictions of the effects of time in Online Supplemental Appendix S2.

Dependent Variables

There are two primary dependent measures: challenging and escalating. As noted above, on every round participants decided whether to *escalate* and also simultaneously decided whether to *challenge* their partner each round. Both outcomes are binary measures where one indicates the participant escalated or challenged their partner, and zero indicates they did not escalate or deferred to their partner. Descriptive statistics for the dependent variables are included in Table 2.

[Table 2 about Here]

Analytic Strategy

Both outcomes are binary, therefore we use logistic regression to test for the effects of the independent variables. Because we analyze 30 rounds of the study for each participant, the data are not independent but instead include multiple observations for each participant. To account for the person-level clustering, we use multilevel models with the participant included as a level-2 random-intercept (see Luke 2004; Rabe-Hesketh and Skrondal 2012 for details). As participants interacted with simulated partners, there is no need to correct for group-level clustering. All tests

of the hypotheses are one-tailed; all other reported tests are two-tailed with significance assessed at the $p < 0.05$ level throughout.

Some of the hypotheses propose a conditional effect, with one effect being moderated by another variable. In linear models, it is common to include an interaction term in a model and examine the interaction coefficient as evidence of moderation. However, in logistic regression models the interaction term cannot be directly interpreted and both the direction and significance level of the interaction term can conflict with the actual nature of the underlying effect (Ai and Norton 2003; Berry, DeMeritt, and Esarey 2010). This difficulty stems from the fact that while logistic regression coefficients are linear in the log-odds, predicted probability curves are non-linear and are bounded between 0 and 1 and thus all eventually converge to the same values. For these reasons, it is recommended to focus on predicted probabilities and to rely on tests of the predicted probabilities in order to determine whether a moderating effect exists across specific levels of interests (see Berry et al. 2010; Long and Freese 2014 for details).⁶ For our purposes, we predict the effect of partner aggression in the two separate experimental conditions (which provide the first differences, or the effect for each specific condition) and then determine whether the effect of partner aggression varies across the two conditions (which is referred to as the second difference; see Berry et al. 2010).

RESULTS

We first present the results and relevant tests of the hypotheses for increasing coercive capability (hypotheses 1, 3), followed by the results for challenging (hypotheses 2, 4-5). Table 3 presents the odds-ratios from the models examining how the cost of escalation and partner aggression influence decisions about escalating and challenging. Figures displaying the effects of condition and partner aggression are available in Online Supplemental Appendix S3.

Escalating

[Table 3 About Here]

As predicted, participants chose to increase their coercive capability less, by a factor of 0.10, when it was high versus medium cost to increase coercive capability ($p < 0.001$; Model 1; see Hypothesis 1). Also supporting our predictions, participants escalate more often when facing partners that were aggressive towards them in their last interaction ($OR = 1.74, p < 0.001$; Model 1; see Hypothesis 3)⁷.

Challenging

Next, we examine the hypotheses for challenging behavior—presented in the right half of Table 3. In support of Hypothesis 2, we find that participants challenge less when escalating is high versus medium cost; increasing cost reduced the odds of challenging by a factor of 0.37 ($p < 0.01$; see Model 3 in Table 2).

Next, we examine how challenging behavior is affected by the direct aggressive behavior of the partner towards the participant. Because we predicted a moderating effect of the cost of increasing coercive capability, we make sets of predictions separately for the medium and high cost conditions. The first differences are the effect of partner aggression given the cost of escalation, for example, whether being challenged by one's partner on a previous round increases or decreases the probability of challenging that partner on the next opportunity. The second difference tests whether the effect of partner aggression varies by the cost of escalation. Table 4 presents the predictions along with tests of the first and second differences.

[Table 4 about Here]

As hypothesis 4 anticipates, we find less challenging against aggressive partners when escalation is expensive. In the high cost condition, the predicted probability of challenging an

aggressive partner is lower than the probability of challenging a non-aggressive partner ($Pr = 0.46$ vs. $Pr = 0.58$, $p < 0.001$). This supports the prediction that aggression has a deterrent effect when the cost of escalation is high. However, we find that partners' aggression does not affect how much individuals challenge when it is medium cost to escalate ($Pr = 0.70$ vs. $Pr = 0.73$, $p = n.s.$). In other words, when individuals can more cheaply escalate, partner aggression no longer serves as a deterrent. Tests of the second difference indicate that partner aggression has a larger deterrent impact when it is costly to escalate compared to when it is medium cost ($p < 0.01$; see Table 4). This provides partial support for hypothesis 5, as we find partner aggression is less of a deterrent in medium compared to high cost conditions—but do not find that this leads to reciprocation of aggression, but instead simply removes the effect of partner aggression.⁸

DISCUSSION

Overall, Study 1 supports most, but not all, of the hypotheses. Consistent with our expectations, when the costs of increasing coercive capability were high (versus medium), participants were less likely to escalate and less likely to challenge their partners (hypotheses 1 & 2). Participants were also more likely to escalate when facing a partner who had directly challenged them on a previous round (hypothesis 3).

Our most novel hypotheses addressed how the cost of escalation moderates the way participants respond to aggressive partners. When the cost was high, participants were likely to defer to partners who had challenged them in the past (supporting hypothesis 4). When escalating was medium cost, individuals were equally likely to challenge partners who had been aggressive versus deferential on the previous round. That is, lowering the cost of escalating eliminated the deterrent effect of aggressive behavior, but did not go so far as reversing it so that aggressive behavior prompted reciprocal aggression (as predicted by hypothesis 5).

STUDY 2

Our theory focuses on understanding aggressive behavior as a function of the cost of escalation. A limitation of Study 1 is that, while we can view the relative effect of moving from a high to a medium cost of escalation, this does not distinguish between the relative and absolute effects of cost. For example, it is possible there is a consistent relative effect of cost, such that higher costs always produce less retaliation than lower costs. Conversely, the effect may vary at different points in the parameter space, such that the effect of cost might differ between high and medium cost compared to medium versus low cost.

In Study 2, we conduct a follow-up study in which we replicate the medium cost condition from study 1, and compare it to a condition in which escalating is low cost. Methodologically, we felt that it was important to replicate the medium cost condition because of the novelty of the findings, whereas the key findings in the high cost condition – that individuals tend to defer to aggressive partners – conceptually replicate the results of several prior experiments (Benard 2013, Benard 2015). Theoretically, this allows us to distinguish between two possibilities: First, lowering the cost from medium to low could further disinhibit retaliation by removing an additional barrier towards aggression, leading to greater retaliation in the low versus medium cost conditions. Second, alternatively, if escalating is inexpensive, individuals may expect their partners to escalate, and so perceive little advantage to challenging aggressive partners. This could lead to lesser retaliation in the low versus medium cost conditions. Practically speaking, this low-cost specification may better align with real-world conditions, if the costs of increasing coercive capability are often not so high that opponents lose whatever social or financial payoffs they seek by doing so. To test these possibilities, we conducted a second study using the same experimental design and drawing a sample from the same

population, but specifying low and medium cost conditions and removing the high cost condition.

Participants

Study 2 includes 91 paid undergraduate volunteers (61 women and 30 men) recruited in the same method as Study 1.⁹ Eight participants were removed from the analysis due to suspicion about some aspect of the study, leaving 83 participants for analysis.

Procedure

Study 2 progressed identically to Study 1 except for changes to the cost of increasing coercive capability. Study 2 had two between-subject conditions: low cost of escalating (10 points) and medium cost (25 points). In the former condition, escalating is 40% of the cost in the latter. This is fairly close to the proportional difference between conditions in study 1 (where the medium cost condition is one third of the high cost), while employing a value that is easy for participants to remember and employ.

RESULTS

Escalating

We again present condensed tables for decisions to escalate and challenge in Table 5; full tables that include the estimates for time block are available in Online Appendix S2. In support of Hypothesis 1, we find that participants escalate more often when it is low cost than when it is medium cost ($OR=5.85, p < 0.001$; Model 1). Based on the main effects shown in Table 5, it appears that participants escalate at similar rates when matched with aggressive and non-aggressive partners ($OR = 1.12, p = ns$; Model 2). However, further analyses examining the moderating effect of cost of escalating suggest that participants in the medium cost condition tend to escalate in response to aggressive partners (replicating Study 1) while those in the low

cost condition are actually *less* likely to escalate. Specifically, we find that individuals are more likely to escalate against aggressive partners when it is medium cost ($Pr = 0.37$ vs. $Pr = 0.26$, first difference $p < 0.001$). In contrast, we find that individuals are *less* likely to escalate against aggressive partners when it is cheap to do so ($Pr = 0.68$ vs. $Pr = 0.76$, first difference $p < 0.05$; second difference $p < 0.001$). This pattern of results is presented in Figure S3.4 in Online Supplemental Appendix S3.

[Table 5 About Here]

Challenging

The right half of Table 5 presents the results for challenging behavior (hypotheses 2, 4-5). Consistent with Study 1, we find that individuals challenge more when escalating is low cost than when it is medium cost; the odds of challenging increase by a factor of 2.95 when escalating is inexpensive ($p < 0.01$). Also consistent with Study 1, we find that individuals challenge less often against partners who challenged them previously than against partners who did not ($OR = 0.58, p < 0.001$).

[Table 6 about Here]

Study 1 showed that direct aggression from partners has a deterrent effect in the high-cost condition (Hypothesis 4), but that when the cost of escalating was medium, direct partner aggression no longer had a deterrent effect (partially supporting hypothesis 5). Study 2 contrasts a low cost condition with the medium cost condition. We use the first and second difference approach as in Study 1 to test for moderation (see Table 6). Consistent with Study 1, partner direct aggression does not increase challenging when it is medium cost to escalate; if anything the trend is in the opposite direction, and is not significant at the .05 level ($Pr = 0.60$ vs. $Pr = 0.55, p = 0.09$, two-tailed test). However, when it is inexpensive to escalate, partner aggression

has a larger and highly significant deterrent effect ($Pr = 0.85$ vs. $Pr = 0.70$, $p < 0.001$). Tests of the second difference suggest that partner aggression has a larger deterrent effect when it is low cost to escalate, compared to situations when it is medium cost ($p < 0.05$).¹⁰

DISCUSSION

In Study 2, we replicated the medium-cost condition from Study 1, and compared it to a condition in which escalating is low cost. The replication of the medium cost condition produced similar results to Study 1. We were especially interested in the comparison of the medium and low cost conditions, and whether participants would be more or less likely to escalate, and to challenge, when faced with partners who had aggressed towards them directly on the prior round.

Interestingly, participants in the low cost condition were *less* likely to challenge partners who directly aggressed against them compared to non-aggressive partners. When compared to the results of Study 1, this suggests a non-linear effect of the cost of escalation on deterrence, where individuals are deterred from challenging partners who have challenged them directly in the past to a greater extent when the cost of escalating is high or low, compared to when it is medium.

To further investigate this apparent non-linear effect, we conducted supplementary analyses more closely examining patterns of aggressive behavior. In these models (see Online Supplementary Appendix S4), we examined participants' willingness to challenge based on the outcome of the prior interaction (i.e., mutual deference, unilateral challenging, and mutual challenging when the participant either wins or loses). For study 1, decreasing the cost of escalating from high to medium does not significantly increase participants' willingness to challenge aggressive partners more than it increases their willingness to challenge deferential partners (i.e., the second differences – the comparisons of the size of the increase in challenging

aggressive versus non-aggressive partners – are not significant). In study 2, in contrast, this analysis suggests that decreasing the cost of escalating from medium to low increased participants' willingness to challenge deferential partners more than it increased their willingness to challenge aggressive partners. The Study 2 finding that individuals defer to aggressive partners more than non-aggressive partners when escalating is low cost, but not when escalating is medium cost, appears to be driven by participants in the low cost condition being more willing to aggress towards apparently non-aggressive partners.

CONCLUSION

Our aim in this paper is to bridge two apparently conflicting lines of research in the study of aggression. Deterrence perspectives argue that aggressive behavior wards off antagonists, by signaling that the aggressor possesses the coercive capability to protect themselves. In contrast, reciprocity perspectives argue that aggressive behavior invites retaliation, by eliciting an emotional, moral, or strategic response from the victim. Given that both perspectives have support, distinguishing the conditions that produce each pattern contributes to understanding the complexity of human aggression. We proposed that a previously-unexamined moderating variable – the cost of escalating – might help explain why individuals respond to aggressors with deference versus retaliation. We expected a relative effect of the cost of escalation, such that partner aggression promotes deference when the cost is higher.

Perhaps our most novel finding is an apparent nonlinear effect of the cost of escalation on responses to aggression. Participants challenge aggressive and non-aggressive partners at similar rates in the medium cost conditions (Studies 1 & 2), but challenged aggressive partners at lower rates than non-aggressive partners when the cost of escalation is relatively high (Study 1) or low (Study 2). Supplementary analyses suggest lowering the cost from medium to low increases

aggression towards deferential partners more than it increases aggression towards aggressive partners, primarily when the participant was previously deferential. Thus, the greater deference shown to aggressive individuals in the low-cost condition appears to be driven by an increase in challenging towards non-aggressive individuals, rather than a decrease in aggression towards aggressive individuals. Most other findings were either in line with our predictions or replicated past work.

In terms of our theory, these findings offer a mix of support and revision. As we expected, the cost of escalation plays a role in explaining when individuals choose retaliation over deference. This highlights the importance of treating coercive capability as a variable rather than a constant, and opens up a potentially important avenue for bridging the deterrence and reciprocity perspectives.

At the same time, the non-linear effect of cost suggests that future work pay attention to both the relative and absolute effects of cost. For some outcomes in our study, there is a straightforward relative effect of cost. For example, escalation increases as the cost of doing so decreases. For other outcomes – such as the deterrent effect of aggression – the absolute costs appear to matter, as the direction of the effect changes when moving from high to medium and medium to low costs.

Although medium cost escalation largely eliminated the deterrent effect of aggression, we do not find evidence for “vengefulness;” participants are not *more* likely to challenge aggressors than non-aggressors. It is possible that our setting – with its relatively narrow focus on material costs and benefits – does not produce sufficient moral outrage, and the overall costs of retaliating in our GHD game are high enough to forestall retaliation. Moralistic retaliation is instrumental for obtaining a number of outcomes including justice or punishment, deterrence,

and self-enhancement or identity restoration. Future work should examine the motivations underlying retaliation, such as signaling, emotions, or morality and justice concerns. For example, one could examine cases in which the participant is harmed by a moral transgression (perhaps using a setting in which a partner beats the participant in an experimental game by cheating versus by playing fairly). Alternatively, it may be that revenge requires some degree of social facilitation or normative pressure (Papachristos 2009), which is not present in our setting.

One important question is the extent to which these findings will generalize to other groups and contexts (Lucas 2003). While our experiments cannot directly answer this question, they do suggest that paying greater attention to the costs – broadly defined – of escalating could offer useful analytic purchase in understanding aggression, and help to evaluate the applicability of our arguments. In particular, social settings vary in actors' capacity or propensity to increase their coercive capability. For instance, fistfights and verbal altercations are common among young men in bars (see Graham et al. 2006), and this is true whether these establishments mainly tailor to a college town crowd or hardened bikers. But in the latter setting, disputants are more likely to deploy weaponry when challenging an adversary. Many reasons could account for these differences across settings– antagonists in middle class bars may perceive greater opportunity costs of employing weapons or this form of escalation might not be a component of their cultural toolkit (Sampson and Bean 2006). Within such settings, actors may also vary in their individual costs of escalation, due to factors such as greater experience with or access to weapons. This may shape the outcome of conflict; for example, actors may be reluctant to engage in conflict when their cost of escalating is high and an antagonist's is low.

At a more macro level, the cost of escalating may help explain conflict among corporate actors, such as organizations or states. Much has been written on the importance of signaling and

credible threats as a means of deterring aggression in international contexts (e.g., Schelling 1960). The costs of escalating vary cross-nationally, depending on factors such as economic development and the threat of sanctions or retaliation for pursuing a weapons program. Such costs may shape a nation's willingness to retaliate when they perceive aggression. The contextual backdrop of hostile encounters affects the perceived physical and emotional costs of escalating and these costs may help to explain aggression and deference.

Broadly speaking, our results also suggest a more general way of approaching deterrence and reciprocity. Instead of asking whether aggression provokes deference *or* retaliation, it may be more productive to disaggregate the considerations – such as strategy, emotions, and morality – that can tip individuals towards one or the other outcome. Conflict is common in social life, and the question of why some conflicts culminate in aggression while others do not is a puzzle worthy of further investigation. The present studies illustrate how the cost of escalating can act as a switch that redirects an individual's tendency to respond to aggression; other such switches likely exist and interact in complicated ways. Uncovering them is likely to both deepen and complicate our understanding of human aggression.

SUPPLEMENTAL MATERIAL

The Online Supplemental Appendix may be found at [insert link here].

ACKNOWLEDGEMENTS

The authors thank Richard Felson and Michael Lovaglia for helpful discussion and feedback on earlier drafts of this manuscript, Lance Bertsch, Zac Childers, Kelsey Cundiff, Josh Doyle, Kristen Forbes, Courtney Hamilton, Lauryn Quick, Kiersten Thaxton, and Lynn Zhang for excellent research assistance, and Bianca Manago for superb lab management.

¹ To illustrate, a study by Jacobs (2000) reports that some drug dealers in St. Louis are less likely to carry weapons in areas with high police presence. This strategy decreases their risk of arrest on a weapons charge, but limits their capacity to retaliate if robbed. In contrast, when police presence (and the corresponding cost of escalating) is lower, drug dealers are more likely to carry weapons, and presumably less deterred by robbery attempts

² We found no main effects of gender (both men and women challenged and increased their coercive capability at equivalent rates), and only one moderating effect of gender. Men challenged deferent partners at higher rates than they challenged aggressive partners ($p < 0.01$) while women challenged deferent and aggressive partners at similar rates ($p = \text{n.s.}$; second difference $p < 0.01$).

³ As an exploratory step, we administered three scales measuring individual-level characteristics related to aggression (endorsement of honor culture values, self-regulation, and past experience with aggression) to respondents at the end of each study. We found little in the way of main or moderating effects of these factors, suggesting our experimental manipulations are responsible for differences observed in the study and not participant characteristics (results available from the authors).

⁴ An alternative approach would be to simulate partners using potentially more realistic assumptions, e.g. that partners who challenge will be more likely to have escalated. This approach requires caution to avoid introducing confounds, e.g. between cost of escalation and partner aggression. Because a strength of experiments is their ability to disentangle variables that are confounded in the real world, we focus on making causal inferences over adding realism. Alternatively, different partner types could be introduced by fully crossing these traits to avoid confounds, as we do for partner reputation (see “Partner reputation”, below). Future work should

examine different partner types or the effects of using real partners in these settings to develop a broader sense of how individuals respond to different partner behaviors. Given that our suspicion rates were relatively low, we do not think our design choices led participants to perceive their partners as unrealistic.

⁵ We presented participants with their partner's reputation for increasing coercive capability, rather than their reputation for challenging, for two reasons. First, the effects of partner's reputation for challenging are documented in past studies (Benard 2013, 2015), and second, there is little reason to escalate but not challenge, and so we expect these two behaviors to be highly correlated (indeed, participants challenged roughly 95% of the time that they escalated).

⁶ While the coefficient on the interaction term should not be directly interpreted, the interaction term itself is needed in the model to allow the underlying effect on the probabilities to vary or not (see Rainey 2015).

⁷ In exploratory analyses, we examined whether the cost of increasing coercive capability moderates the effect of partner aggression on the decision to increase coercive capability. We find that participants were more likely to escalate against aggressive partners both when it was medium or high cost (both first differences $p < 0.001$), and that the effect of partner aggression on increasing coercive capability was similar regardless of the cost (second difference $p = \text{n.s.}$). Figure A1 in Appendix A illustrates this pattern of results.

⁸ We conducted additional analyses to examine the effect of partners' reputation for escalating, i.e., the number of times the partner had chosen to escalate in the past 5 rounds. A reputation for escalating did not significantly increase participants' willingness to escalate. Consistent with past work (Benard 2013, 2015), participants were less likely to challenge individuals with a reputation for escalating on prior rounds; this tendency was not significantly moderated by the

cost of escalation. These results are reported in Online Supplementary Appendix S1: Effect of Partner Reputation.

⁹ No main effects of gender were predicted or found; one moderating effect of gender was found.(see also, footnote 2). Men are more likely to escalate when facing aggressive versus deferent partners ($p < 0.05$), whereas women escalate at similar rates regardless of partner aggression ($p = \text{n.s.}$; second difference $p < 0.05$).

¹⁰ As in Study 1, we examined the effect of partners' reputation for escalating on participants' willingness to escalate and challenge. Similar to Study 1, we found that partner reputation did not significantly affect participants' decision to escalate, but did reduce participants' willingness to challenge their partner. As in Study 1, the deterrent effect of partner reputation was not moderated by the cost of escalation. See Online Supplementary Appendix S1 for details.

REFERENCES

- Ai, Chunrong and Edward C Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80(1):123-29.
- Anderson, Elijah. 2000. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*: WW Norton & Company.
- Aquino, Karl, Thomas M Tripp and Robert J Bies. 2001. "How Employees Respond to Personal Offense: The Effects of Blame Attribution, Victim Status, and Offender Status on Revenge and Reconciliation in the Workplace." *Journal of Applied Psychology* 86(1):52.
- Ariely, Dan and George Loewenstein. 2006. "The Heat of the Moment: The Effect of Sexual Arousal on Sexual Decision Making." *Journal of Behavioral Decision Making* 19(2):87-98.
- Beckerman, Stephen, Pamela I Erickson, James Yost, Jhanira Regalado, Lilia Jaramillo, Corey Sparks, Moises Iromenga and Kathryn Long. 2009. "Life Histories, Blood Revenge, and Reproductive Success among the Waorani of Ecuador." *Proceedings of the National Academy of Sciences* 106(20):8134-39.
- Benard, Stephen. 2013. "Reputation Systems, Aggression, and Deterrence in Social Interaction." *Social Science Research* 42(1):230-45.
- Benard, Stephen. 2015. "The Value of Vengefulness: Reputational Incentives for Initiating Versus Reciprocating Aggression." *Rationality and Society*:1043463115576135.
- Berg, Mark T and Rolf Loeber. 2015. "Violent Conduct and Victimization Risk in the Urban Illicit Drug Economy: A Prospective Examination." *Justice quarterly* 32(1):32-55.
- Berry, William D, Jacqueline HR DeMeritt and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?". *American Journal of Political Science* 54(1):248-66.
- Blau, Peter Michael. 1964. *Exchange and Power in Social Life*: Transaction Publishers.
- Bowles, Samuel and Herbert Gintis. 1993. "The Revenge of Homo Economicus: Contested Exchange and the Revival of Political Economy." *The Journal of Economic Perspectives* 7(1):83-102.
- Cheah, Ban Chuan. 2009. "Clustering Standard Errors or Modeling Multilevel Data." *University of Columbia*:2-4.
- Clark, Bruce H and David B Montgomery. 1998. "Deterrence, Reputations, and Competitive Cognition." *Management Science* 44(1):62-82.
- Cooney, Mark. 1998. *Warriors and Peacemakers: How Third Parties Shape Violence*: NYU Press.
- Crowley, Philip H. 2000. "Hawks, Doves, and Mixed-Symmetry Games." *Journal of Theoretical Biology* 204(4):543-63.
- De Mesquita, Bruce Bueno, James D Morrow and Ethan R Zorick. 1997. "Capabilities, Perception, and Escalation." *American Political Science Review* 91(01):15-27.
- Faris, Robert and Susan Ennett. 2012. "Adolescent Aggression: The Role of Peer Group Status Motives, Peer Aggression, and Group Characteristics." *Social Networks* 34(4):371-78.
- Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415(6868):137-40.
- Felson, Richard B., Mark T. Berg, and Meghan L. Rogers. 2014. "Bring a Gun to a Gunfight: Armed Adversaries and Violence Across Nations." *Social Science Research* 47: 79-90.
- Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10(2):171-78.

- Frank, Robert H. 1988. *Passions within Reason: The Strategic Role of the Emotions*: WW Norton & Co.
- Gherardi, Francesca. 2006. "Fighting Behavior in Hermit Crabs: The Combined Effect of Resource-Holding Potential and Resource Value in *Pagurus Longicarpus*." *Behavioral Ecology and Sociobiology* 59(4):500-10.
- Gould, Roger V. 2003. *Collision of Wills: How Ambiguity About Social Rank Breeds Conflict*: University of Chicago Press.
- Halevy, Nir, Lilach Sagiv, Sonia Roccas and Gary Bornstein. 2006. "Perceiving Intergroup Conflict: From Game Models to Mental Templates." *Personality and Social Psychology Bulletin* 32(12):1674-89.
- Helm, Bob, Thomas V Bonoma and James T Tedeschi. 1972. "Reciprocity for Harm Done." *The Journal of social psychology* 87(1):89-98.
- Jacobs, Bruce A. 2000. *Robbing Drug Dealers: Violence Beyond the Law*: Transaction Publishers.
- Jacobs, Bruce A and Richard Wright. 2006. *Street Justice: Retaliation in the Criminal Underworld*: Cambridge University Press.
- Johnstone, Rufus A and Redouan Bshary. 2004. "Evolution of Spite through Indirect Reciprocity." *Proceedings of the Royal Society of London B: Biological Sciences* 271(1551):1917-22.
- Lawler, Edward J, Rebecca S Ford and Mary A Blegen. 1988. "Coercive Capability in Conflict: A Test of Bilateral Deterrence Versus Conflict Spiral Theory." *Social Psychology Quarterly*:93-107.
- Loch, Christoph, Michael Yaziji and Christian Langen. 2001. "The Fight for the Alpha Position: Channeling Status Competition in Organizations." *European Management Journal* 19(1):16-25.
- Long, J Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*: Stata press.
- Lucas, Jeffrey W. 2003. "Theory-Testing, Generalization, and the Problem of External Validity." *Sociological Theory* 21(3):236-53.
- Luckenbill, David F. 1977. "Criminal Homicide as a Situated Transaction." *Social Problems* 25(2):176-86.
- Luke, Douglas A. 2004. *Multilevel Modeling*, Vol. 143: Sage.
- Molm, Linda D. 1997. *Coercive Power in Social Exchange*: Cambridge University Press.
- Molm, Linda D. 2010. "The Structure of Reciprocity." *Social Psychology Quarterly* 73(2):119-31.
- Nisbett, Richard E and Dov Cohen. 1996. *Culture of Honor: The Psychology of Violence in the South*: Westview Press.
- Papachristos, Andrew V. 2009. "Murder by Structure: Dominance Relations and the Social Structure of Gang Homicide." *American journal of sociology* 115(1):74-128.
- Podolny, Joel M and Fiona M Scott Morton. 1999. "Social Status, Entry and Predation: The Case of British Shipping Cartels 1879–1929." *The Journal of Industrial Economics* 47(1):41-67.
- Rabe-Hesketh, Sophia and Anders Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*: STATA press.
- Rainey, Carlisle. 2016. "Compression and Conditional Effects: A Product Term Is Essential When Using Logistic Regression to Test for Interaction." *Political Science Research and Methods* 4(03):621-39.
- Sampson, Robert J and Janet L Lauritsen. 1994. "Violent Victimization and Offending: Individual-, Situational-, and Community-Level Risk Factors." *Understanding and preventing violence* 3.
- Schelling, Thomas C. 1980. *The Strategy of Conflict*: Harvard university press.
- Snyder, Glenn H. 1971. "'Prisoner's Dilemma' and 'Chicken' Models in International Politics." *International Studies Quarterly* 15(1):66-103.
- Tedeschi, James T and Richard B Felson. 1994. *Violence, Aggression, and Coercive Actions*: American

- Psychological Association.
- Topalli, Volkan, Richard Wright and Robert Fornango. 2002. "Drug Dealers, Robbery and Retaliation. Vulnerability, Deterrence and the Contagion of Violence." *British Journal of Criminology* 42(2):337-51.
- Yamagishi, Toshio. 1986. "The Provision of a Sanctioning System as a Public Good." *Journal of Personality and Social Psychology* 51(1):110.

TABLES

Table 1: The Generalized Hawk-Dove Game.

		<u>Partner's Choice</u>	
		Challenge	Defer
<u>Own Choice</u>	Challenge	$p(V-W)-(1-p)L = \lambda p-L$	V, 0
	Defer	0, V	V/2-T

Payoff Table Shown to Participants in Study 1

		<u>Partner's Choice</u>	
		Challenge	Defer
<u>Own Choice</u>	Challenge	Win: earn 50 points Lose: lose 100 points	You: earn 50 points Partner: earns 0 point
	Defer	You: earn 0 points Partner: earn 50 points	You: earn 15 points Partner: earns 15 points

Notes: Parameters: p = probability of winning, V = value of resource pool, W = cost of winning, L = cost of losing. For simplicity, $\lambda = V+L-W$.

Table 2: Descriptive Statistics for Dependent Variables

Variable	Proportion	Min	Max
Study 1			
Escalating	0.28	0	1
Challenge	0.61	0	1
Study 2			
Escalating	0.51	0	1
Challenge	0.65	0	1

Notes: (1) Study 1: $N_{\text{level-1}} = 3,241$; $N_{\text{level-2}} = 109$ **(2)** Study 2 ($N_{\text{level-1}} = 2,490$; $N_{\text{level-2}} = 83$)

Table 3: Random-Intercept Multilevel Logistic Regression Results for Escalating and Challenging by Cost and Partner Aggression — Study 1 ($N_{\text{level-1}} = 3,241$; $N_{\text{level-2}} = 109$)

	<u>Escalating</u>		<u>Challenging</u>	
	1	2	3	4
High Cost	0.10***	0.08***	0.37**	0.47*
.....	(0.02)	(0.02)	(0.12)	(0.16)
Partner Aggression	1.74***	1.56***	0.70***	0.87
.....	(0.17)	(0.17)	(0.06)	(0.11)
High Cost X Partner Aggression..		1.50		0.63*
.....		(0.33)		(0.11)
Partner Reputation	1.00	1.00	0.61***	0.61***
.....	(0.03)	(0.03)	(0.02)	(0.02)
Constant	0.73	0.77	17.52***	15.76***
.....	(0.15)	(0.16)	(4.75)	(4.32)

Notes: (1) Odds-ratios are reported, (2) standard errors in parentheses, (3) a level-2 random-intercept for the participant is included to account for clustering, (4) binary indicator variables for time block of the study are included in all models but not shown in the table, (5) * $p < .05$, ** $p < .01$, *** $p < .001$

Table 4: Predictions and Tests of Second Differences from Random-Intercept Multilevel Models Predicting Challenging Behavior — Study 1 ($N_{\text{level-1}} = 3,241$; $N_{\text{level-2}} = 109$)

	Pr(Challenging)	Test of 1 st Difference	Test of 2 nd Difference
<u>Medium Cost</u>			
Not Aggressive	0.73	0.73 – 0.70 =	0.02 – 0.13 =
	(0.04)	0.02	-0.10**
Aggressive	0.70		
	(0.04)		
<u>High Cost</u>			
Not Aggressive	0.58	0.58 – 0.46 =	
	(0.05)	0.13***	
Aggressive	0.46		
	(0.05)		

Notes: (1) Probabilities of challenging are presented, (2) standard errors in parentheses, (3) a level-2 random-intercept for the participant is included to account for clustering, (4) all models include controls for reputation of partner and time block of the experiment, (5) * $p < .05$, ** $p < .01$, *** $p < .001$

Table 5: Random-Intercept Multilevel Logistic Regression Results for Escalating and Challenging by Cost and Partner Aggression — Study 2 ($N_{\text{level-1}} = 2,490$; $N_{\text{level-2}} = 83$)

	<u>Escalating</u>		<u>Challenging</u>	
	1	2	3	4
Low Cost	5.85***	9.08***	2.95**	4.39***
.....	(2.26)	(3.66)	(1.19)	(1.85)
Partner Aggression	1.12	1.65***	0.58***	0.79
.....	(0.11)	(0.22)	(0.06)	(0.11)
Low Cost X		0.42***		0.47***
Partner Aggression..		(0.08)		(0.10)
Partner Reputation	0.96	0.96	0.67***	0.66***
.....	(0.03)	(0.03)	(0.02)	(0.02)
Constant	0.66	0.54*	7.39***	6.36***
.....	(0.19)	(0.16)	(2.31)	(2.01)

Notes: (1) Odds-ratios are reported, (2) standard errors in parentheses, (3) a level-2 random-intercept for the participant is included to account for clustering, (4) binary indicator variables for time block of the study are included in all models but not shown in the table, (5) * $p < .05$, ** $p < .01$, *** $p < .001$

Table 6: Predictions and Tests of Second Differences from Random-Intercept Multilevel Models Predicting Challenging Behavior — Study 2 ($N_{\text{level-1}} = 2,490$; $N_{\text{level-2}} = 83$)

	Pr(Challenging)	Test of 1 st Difference	Test of 2 nd Difference
<u>Low Cost</u>			
Not Aggressive	0.85	0.85 – 0.70 =	0.15 – 0.05 =
	(0.04)	0.15***	0.10*
Aggressive	0.70		
	(0.06)		
<u>Medium Cost</u>			
Not Aggressive	0.60	0.60 – 0.55 =	
	(0.06)	0.05	
Aggressive	0.55		
	(0.06)		

Notes: (1) Probabilities of challenging are presented, (2) standard errors in parentheses, (3) a level-2 random-intercept for the participant is included to account for clustering, (4) all models include controls for reputation of partner and time block of the experiment, (5) * $p < .05$, ** $p < .01$, *** $p < .001$