Prospect Theory, Fairness, and the Escalation of Conflict at Negotiation Impasse

Miettinen, Topi; Ropponen, Olli; Sääskilahti, Pekka

Published in:
The Scandinavian Journal of Economics

DOI:
10.1111/sjoe.12384

Publication date:
2019

Document Version
Peer reviewed version, als known as post-print

Prospect theory, fairness, and the escalation of conflict at negotiation impasse*

Topi Miettinen †
Hanken School of Economics and Helsinki GSE

Olli Ropponen ‡
VATT Institute for Economic Research

Pekka Sääskilahti §
Compass Lexecon

Abstract

We study a bilateral negotiation setup where at bargaining impasse the disadvantaged party chooses whether to escalate the conflict or not. Escalation is costly for both parties and it results in a random draw of the winner of the escalated conflict. We derive the behavioral predictions of a simple social utility function which is convex in disadvantageous inequality, thus connecting the inequity aversion and the prospect theory models. Our causal laboratory evidence is to a large extent consistent with the predicted effects. Among other things, the model correctly predicts that the escalation rate is higher when escalation outcomes are riskier and the disagreement rate is lower when the cost of escalating the conflict is higher.

KEYWORDS: bargaining; conflict; inequity aversion; loss aversion; quantal response equilibrium

JEL CODES: C72, C91, D03

1 Introduction

Wage negotiations between a corporate employer and a labor union may stall and escalate into a strike which harms both parties. Settlement negotiations between a plaintiff and a defendant of a lawsuit may stall and

*Views presented do not necessarily represent those of our employers. We thank C.Göring, T.Mäkelä, and M.Ploner for research assistance and G.Bolton, W.Güth, A.Isoni, K.Kotakorpi, M.Liski, P.H.Matthews, T.Nurminen, M.Ploner, O.Rydval, A.Lindholm, seminar audiences at NHH, Helsinki GSE, IFN, Innsbruck, Jena, EEA-ESEM, and Bilkent University for insightful comments. We acknowledge the financial support of Norwegian Research Council (250506), IPR University Center, Yrjö Jahnsson Foundation, and Max Planck Institute of Economics.

†*Address:* Hanken School of Economics, Arkadiankatu 7, PO Box 479, Fi-00101 Helsinki, *E-mail:* topi.miettinen@hanken.fi

‡*Address:* VATT Institute for Economic Research, Arkadiankatu 7, PO Box 1279, Fi-00101 Helsinki, *E-mail:* olli.ropponen@vatt.fi

§*Address:* Compass Lexecon, Aleksanterinkatu 15B, Fi-00100 Helsinki, Finland, E-mail: psaaskilahti@compasslexecon.com

the plaintiff may take the case to court implying high legal expenses for both parties and uncertainties about the final verdict. Negotiations in a territorial conflict may stall and trigger armed conflict with devastating consequences. There are more than 10 million battle casualties across the globe since the second world war (Lacina and Gleditsch, 2005); in year 2000 in U.S State courts alone, about 20 million cases were filed of which about 3-4% end up in trial leaving the courts with a work-load of about million cases yearly (Ostrom et al., 2003); strikes and labor unrest have a negative impact on productivity and product quality (Kruger and Mas, 2004; Mas, 2008) and Gruber and Kleiner (2012) show that nurses' strikes increased in-hospital mortality by 18.3 percent in the state of New York. The failure of bargaining is a key prerequisite for all of these inefficiencies to arise.[1]

In this paper, we design a simple non-framed experiment to better understand how inefficient conflict comes about between two individuals, and how the disadvantaged dispute party may engage in escalation of conflict when negotiations stall. We are interested in testing the implications of a social utility model that makes explicit the connection between the prospect theory value function (Kahneman and Tversky, 1979, 1992) and the inequity aversion model of Fehr and Schmidt (1999). We hypothesize that at a negotiation impasse, social comparison importantly influences the decision whether to escalate conflict or not and that the convexity of the social utility in disadvantageous payoff inequality is the key to understand these decisions. These hypotheses are captured in an "inequity-as-loss" utility function that we propose. This form of the social utility function was suggested by Loewenstein et al. (1989), who found that disadvantageous inequality can be accounted as a loss in the prospect theory sense. The successful social preference literature that followed (Bolton, 1991; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002) paid less attention to the curvature properties and focused more on first-order effects of inequity aversion and fairness. If disadvantageous inequality is perceived as a loss in this manner, then the implications could be dramatic for settlement patterns: the disadvantaged provokers will escalate conflict more the riskier the escalation outcomes are - not less as suggested by risk aversion.[2]

From an applied perspective, there is a pressing need for understanding such effects due to their potential effects on such important frictions as labor disputes or the burden of courts. More generally, better joint models of risk and other-regarding preference hold a promise of yielding higher explanatory power in any strategic interaction context.

In our experimental design, parties first attempt a settlement. A failure to strike a deal puts one of the

[1] The rational explanations of strikes and industrial conflict (Kennan and Wilson, 1989), of failed pretrial settlement (Spier, 2007), and of armed conflict (Jackson and Morelli, 2011) are increasingly well understood.

[2] In addition to social comparison, high loss references may be driven for instance by high aspirations set at the negotiation table (Korobkin, 2002,Karagözoglu and Keskin 2018). See also Cox et al. (2007) and Bolton and Ockenfels (2000) for models of non-linearities in inequity aversion. Eisenkopf and Teyssier (2013) study the effect of envy and loss aversion in contests but without convexity effects.

parties at a disadvantaged position in the sense that her pecuniary payoff is lower than that of the opposing side at all ensuing conflict outcomes. This underdog is given an option either to acquiesce or to engage in inefficient rent-seeking, i.e. to escalate conflict in order to potentially reach a still disadvantaged but less unequal outcome. The decision to escalate results in a lottery with an exogenous and publicly known probability of winning and losing, and equally large publicly known expenses to each side of the dispute. We experimentally vary (i) the probability of winning of the disadvantaged party, (ii) the cost of escalation of conflict, (iii) and while preserving the expected payoffs at escalation, whether the escalation outcome is risky or certain. In our setup a lower probability of winning and higher escalation costs are perfect substitutes in reducing escalation incentives for a risk-neutral and self-interested underdog.

Regarding escalation, we observe that (i) a lower escalator's winning probability and (ii) higher costs of escalation both reduce the escalation rate. We also observe, going against risk-aversion, that (iii) greater variation in escalation outcomes increases the escalation rate. Among the conditions where rational self-interest predicts no escalation, the observed escalation rate is highest when escalation is costly and offers a high chance of rendering payoffs more equal. In fact with risky outcomes, escalation is more frequent than refraining from it even if doing so is suboptimal from a self-interested perspective. Regarding negotiations preceding escalation choices, we find that settlement rates are highest when conflict escalation is expensive. With high costs, escalation threat makes the negotiators more careful in seeking advantage in settlement.

The observed escalation patterns are to a large extent in line with the predictions of the proposed inequity-as-loss model. The model even explains why lowering the underdog's probability of winning and making escalation outcomes less risky curbs inefficient escalation more effectively than increasing the costs of escalation and making the outcomes more risky.[3]

Moreover, when embedded in a logit quantal response equilibrium framework (McKelvey and Palfrey, 1998; Goeree et al. 2016), the comparative statics predictions capture well virtually all the treatment effects on the observed settlement rates and the escalation rates. This is in contrast with the predictions of the self-interested risk-neutral subgame perfect Nash equilibrium which fails to pass the hurdle. Thus, our design allows to point out some limitations of prescriptive rationality assumptions in empirical work which can be circumvented by the adoption of more descriptive theoretical concepts.

Our evidence is consistent with the idea that perceived unfairness of settlement impasse triggers loss-perception in the disadvantaged party. The diminishing marginal sensitivity to losses and the implied preference for risk in attempting reconciliation[4] result in socially inefficient escalation of conflict. Bellemare et

[3]Robson (1992) theoretically studies the effect of status concerns on risk-taking and shows that utility may become convex in wealth due to indirect wealth effectsif the lottery provides an opportunity to surpass at least one other individual in wealth ranking. In our experiment the disadvantaged party earns less at all conflict outcomes.

[4]See Laury and Holt (2008) for further evidence and discussion of diminishing marginal sensitivity or the so called reflection effect and risk preference elicitation.

al. (2008) provide evidence that such a reflection effect among the disadvantaged parties plays a role in explaining bargaining outcomes. Our paper complements their results by explicitly focusing on the effects of convex utility on conflict escalation decisions at the bargaining impasse rather than merely on negotiation behavior. We point out that conflict escalation tendencies are particularly prevalent when outcomes are risky and there is an opportunity for reducing the payoff inequality, thus underscoring the role of the convexity of the utility in payoff-inequality. However the anticipation of escalation may partially remedy the inefficiencies by influencing settlement rate in the negotiation table.

An experimental literature on the interaction of risk and social preferences is only emerging and there is no shared understanding of how to best model such effects (Trautmann and Vieider, 2012). Evidence from Brennan et al. (2008), Bault et al. (2008), Haisley et al. (2008), Bolton and Ockenfels (2010), Linde and Sonnemans (2012), Rohde and Rohde (2011), López-Vargas (2014), Andersson et al. (2015), and Gamba et al. (2016) suggests that redistributive decision making under risk depends heavily on the context and auxiliary design attributes (Guala, 2005; Lowenstein, 1989; List, 2007; Bardsley, 2008).

Compared to the above contributions with a passive recipient, we focus on a case where a party facing disadvantageous inequality[5] has an option to choose a costly redistributive gamble after failed negotiations. Charness and Rabin (2002) have pointed out how other-regarding concerns are responsive to such contextual triggers and how the disadvantage aversion model that we also utilize serves particularly well as a simple motivational model in these cases (see also Bolton, 1991). The novel feature we introduce to the analysis of conflict escalation at negotiation impasse is the convexity of the utility function in the social loss domain.

Our study also relates to the experimental literature on bargaining in the shadow of conflict that examines, among other things, the effect of the asymmetry of conflict on bargaining outcomes (Hoffman and Spitzer, 1985; Kimbrough and Shremeta, 2014; Kimbrough et al., 2014; Herbst et al., 2017; Dechenaux et al., 2015). Anbarci and Feltovich (2013) find that the negotiation strategies do not react to conflict asymmetries as much as the selfish sequentially rational theory would predict, but that quantal response equilibrium and other-regarding preferences can account for the observed patterns. As opposed to a typical contest game, we abstract from strategic uncertainty in conflict escalation by allowing only the disadvantaged party to escalate conflict and by imposing exogenous and publicly known probabilities of winning and losing. Moreover, the private returns from conflict escalation are negative in our main treatments where rational self-interest explanations would predict no escalation (Konrad, 2009).

In the next section, we lay out the model and the experimental setup. The inequity-as-loss model is introduced in Section 3 and the behavioral predictions are derived. In Section 4.1 the empirical results regarding conflict escalation behavior are studied, and Section 4.2 deals with the behavioral patterns in

[5] Trautman and Vieider (2012) coin this the social loss domain.

settlement negotiations. We discuss the results in Section 5.

2 Experimental design

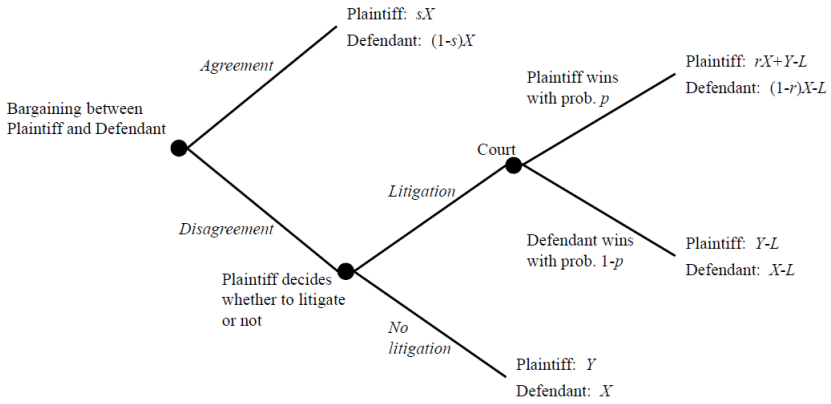

Figure 1: Negotiation game tree

2.1 Framework and experiment setup

In our experimental design, two parties, the provoker (P) and the defender (D), first engage in settlement negotiations over the sharing of value X. A failure to strike a deal puts P at a disadvantaged position in the sense that her pecuniary payoff is lower than that of D at all outcomes – all P payoffs ensuing a bargaining impasse fall into the social loss domain and thus neither loss-aversion nor curvature in the social gains domain confounds identification. At the impasse, the underdog P is given an option either to acquiesce or to engage in inefficient rent-seeking, i.e. to escalate conflict, which is resolved through a lottery with an exogenous and publicly known probability of winning and losing, and equally large publicly known expenses to each side.

Formally, if the parties reach a negotiated agreement, they share the value X in corresponding shares. Let us denote P's share in such an agreement by s (so that P gets sX) and D's share is thus $1 - s$. In the experiment we set $X = 200$. If the negotiations break without an agreement, P will have a possibility to escalate conflict to claim a share of X. Escalation is costly as each party incurs an escalation cost L, identical for both parties. If P wins the escalated conflict, he receives rX where $r = 0.4$ in the experiment. The probability that P wins is p, which is public information. Thus if P decides to escalate, then his expected monetary payoff is

$$\Pi_P = prX + Y - L, \tag{1}$$

$r = 0.4$, $Y = 10$	BASE $p = 0.7,\ \ L = 10$	HIGH $p = 0.7,\ \ L = 58$	LOW $p = 0.1,\ \ L = 10$
Risky outcomes (P win, p) (D win, $1-p$)	$\pi_P = 80, \quad \pi_D = 110$ $\pi_P = 0, \quad \pi_D = 190$	$\pi_P = 32, \quad \pi_D = 62$ $\pi_P = -48, \quad \pi_D = 142$	$\pi_P = 80, \quad \pi_D = 110$ $\pi_P = 0, \quad \pi_D = 190$
Certain outcomes	$\Pi_P = 56, \quad \Pi_D = 134$	$\Pi_P = 8, \quad \Pi_D = 86$	$\Pi_P = 8, \quad \Pi_D = 182$

Table 1: Conflict escalation payoffs across conditions.

and the expected monetary payoff of D is

$$\Pi_D = (1 - pr)\, X - L. \tag{2}$$

If P does not escalate, then P gets Y while D keeps the entire value X. In the experiment we set $Y = 10$. With sequentially rational self-interest, Y has no impact on the optimality of escalation, yet impasse becomes the only rational bargaining outcome in certain circumstances (see Section 3.1). The game is illustrated in Figure 1.

We have three parameter conditions in the experiment: *BASE* where P's probability of winning the escalated conflict is relatively high and the costs of escalation are relatively low ($p = 0.7$, $L = 10$), *LOW* where P's probability of winning is reduced while costs remain low at the BASE level ($p = 0.1$, $L = 10$), and *HIGH* where P's winning probability is maintained at the high BASE level but costs of escalation (for both parties) are increased to a high level ($p = 0.7, L = 58$). Notice that in LOW, both the winning probability of the provoker and the costs of escalation are lower than in HIGH. The parameters are chosen so that the expected payoff for P (but not for D) coincides in HIGH and LOW.

We consider deterministic (CERTAIN) and stochastic (RISKY) escalation outcomes. The deterministic escalation outcomes differ from the stochastic only in that the former implement the expected conflict escalation payoffs of both parties with certainty whereas the stochastic escalation outcomes truly implement a random draw using the publicly known P's probability of winning. Thus the lottery in the case of stochastic escalation outcome is a mean-preserving spread of the provoker's payoff in the deterministic escalation outcome case from the perspective of the provoker's private returns. The escalation payoffs in the experiment are given in Table 1.

In the experiment, for the sake of tractability, negotiations take a specific form where each party makes a take-it-or-leave-it offer to the other and one of the players is (ex-post) randomly drawn as the actual proposer, each with probability 50%. In this special case of random-proposer ultimatum bargaining: the randomly drawn proposer has all bargaining power in sketching a proposal and the other party is only granted a right to veto it. Asking for each party to contrive a proposal for one contingency and a *minimal acceptable offer (MAO)* for the other within a match allows us to collect more informative negotiation plans

in a concise and simple manner.

2.2 Execution of the experiment

The computerized experiment was conducted in the laboratory of the Max Planck Institute of Economics in Jena. Participants were 316 undergraduates from the University of Jena, from different fields of study. Participants were recruited using the ORSEE software (Greiner, 2015) and the experiment was programmed with the z-Tree software (Fischbacher, 2007).

At the beginning of each session, participants were seated at visually isolated computer terminals where they received a hardcopy of the German instructions. The experiment started after all participants had successfully completed a control questionnaire ensuring their understanding.[6] At the beginning of each session, each subject was randomly assigned one of the two roles (P or D) and one of the matching groups or *subsessions* of a session (RISKY or CERTAIN). One quarter of the participants was randomly assigned to each of the four role-subsession constellations. The instructions and the control questions are presented in the online appendix and the decision screens are available upon request.

Each experimental session lasted for 8 rounds; each P (D) played once against each D (P) in subsession RISKY, and likewise for subsession CERTAIN. Once all rounds had been played, the outcome of one round was randomly drawn for the actual payment. Each round consisted of the game illustrated in Figure 1. We used the strategy vector method in eliciting the choices so that each negotiator (each P and D) chose her proposal and MAO without knowing whether the randomly drawn proposer is P or D, and P also chose whether to escalate conflict or not without knowing whether an agreement will be reached at the negotiation stage. To keep the design simple and not to overburden the subjects, we chose not to condition the escalation choice on who was assigned the proposer (responder) role in the negotiation stage.[7] The opponent's choices (but not the random draws of nature) were revealed at the end of each repetition of the game.

The earnings of the experiment were presented in experimental currency units (ECU) with 1 ECU = 0.07 euro. Each P could make losses in any given round including the round randomly drawn for payment. The incurred losses were subtracted of the show-up fee of 3.5 euros which was announced in the opening paragraph of the experimental instructions. Thus the aggregate payment to a subject was never negative. The average earnings were 11.50 euros. The average duration of a session was 1 hour and 20 minutes.

Once the negotiation and escalation choices were elicited, we asked each subject to guess the choices

[6] If a participant could not answer a control question, we did not allow her to proceed to the actual experiment until understanding was ensured. By raising a hand, a subject could ask a laboratory operator to come to her cabin and the subject could pose further questions to the operator individually. About 5% of the subjects posed further questions regarding the instructions, and eventually none of the subjects were excluded from the experiment.

[7] The model that we test in this paper is a consequentialist outcome-based model. In such model, this simplification does not matter.

made by the agent on the opposing side. These guesses were incentivized. Each correct guess yielded a supplementary payoff of 11 ECU. A payoff of 1 ECU was subtracted for each unit (ECU) by which the subject's guess missed the actual negotiation choice so that missing the actual choice (proposal or acceptance threshold) by 10 units delivered 1 ECU and missing by a larger margin than that gave no supplementary payoff at all. To incentivize the binary escalation choice, we used the proper scoring rule which we discretized to simplify exposition.[8] Each D could thus pick one of the following five guesses: that P surely escalates (refrains from escalating), that P is more likely to escalate (to refrain from escalating), and that escalation of conflict and refraining from it are equally likely. In the end of the experiment, one of the guesses was randomly drawn for payment from all rounds but for the round whose negotiation and escalation choices were paid for. Once beliefs were elicited the actual strategy of the opponent was revealed to the subject and she was also reminded of her own strategy. Thus the participants did not learn any population statistics about escalation or negotiation choices nor the outcome of the escalated conflict between two periods of interaction. This left room for learning only from private experiences. The experiment then proceeded to the following round where each participant was matched with a new subject in the opposing role (perfect strangers) thus undermining any repeated game or reputation incentives.

In addition to the random experimental variation between RISKY or CERTAIN and whether one played P or D (where the assignment to RISKY-P, CERTAIN-P, RISKY-D, CERTAIN-D was equally likely as explained above), there was exogenous variation as to whether BASE, HIGH or LOW condition was applied. Each session consisted of 8 rounds. In each session, there were three treatment blocks. Each block consisted of consecutive rounds during which the cost of escalation and P's probability of winning were held constant at levels of BASE, HIGH or LOW conditions (as described in the previous subsection and Table 1). A BASE block lasted for 2 rounds, a HIGH block lasted for 3 rounds, as did also a LOW block. All subjects of a subsession at a given round played the same condition.

The order of blocks was varied from session to session. Subsessions with risky escalation outcomes (each with 8 provokers and 8 defenders) had the following variation in the order of blocks: 1 subsession with order BASE/HIGH/LOW, 1 subsession with order BASE/LOW/HIGH, 2 subsessions with HIGH/BASE/LOW, HIGH/LOW/BASE, LOW/HIGH/BASE, and LOW/BASE/HIGH each, i.e. altogether 10 subsessions with rRISKY. In an analogous manner and with the same orderings of blocks, there were altogether 10 subsessions with CERTAIN. In addition to the comparison between RISKY and CERTAIN, the key comparison of interest is between the HIGH and the LOW conditions. Since the experimental identification is cleanest when comparing first round behavior, we wanted to double the number of orderings with constellations starting with HIGH and LOW, and thus there is only one risky outcomes subsession and one certain outcomes

[8] See Schlag et al. (2015) for a survey of belief elicitation methods in experimental economics.

subsession with each of the two orders starting with the BASE.[9]

To summarize, the variation in RISKY/CERTAIN was between subjects only, the variation in BASE/HIGH/LOW was both within and between subjects. Moreover 16 of the subjects in each session were randomly allocated to role P and the remaining of 16 subjects to role D, each playing in a fixed role over the eight rounds with RISKY or CERTAIN fixed for the eight rounds, once against each of the subjects in the opposing role. There were 6 treatment conditions (see Table 1).[10]

3 Theoretical predictions

The key feature in our setup is that, when there is bargaining impasse, the underdog, i.e. the provoker, earns a lower payoff than the opposing party, the defender, at all outcomes whether the underdog escalates conflict or not. An agreement on a fifty-fifty split in the negotiations is the only way of reaching equal payoffs. At the escalation stage, that opportunity has already been lost.

Our primary hypothesis is that under those circumstances, the key factor that drives the provoker's behavior is social comparison and the sensation of disadvantageous inequality. We formalize this with a prospect theory value function of the following form and call it the *behavioral value function* in the sequel

$$v_p = \pi_P - \lambda(\pi_D - \pi_P)^\gamma, \tag{3}$$

where $0 < \gamma < 1$ due to diminishing marginal sensitivity (e.g. the so called reflection effect). Variables π_P and π_D indicate the monetary payoff for the provoker and defender, respectively. The defender's payoff constitutes the gain-loss reference, so that the provoker experiences a payoff lower than that of the defender as a loss. It is easy to see that this value function is increasing and convex in π_P thus capturing the essentials of a prospect theory value function in the loss domain (Kahneman and Tversky, 1979, 1992). The parameter λ is the loss aversion parameter.[11]

It is equally straightforward to notice that the value function is also closely related to the inequity aversion model of Fehr and Schmidt (1999). The cost of disadvantageous inequality in their model would be written in the form $-\lambda(\pi_D - \pi_P)^\gamma$ where $\gamma = 1$ and parameter λ describes the aversion to disadvantageous inequality. The only modification to the original model is to allow for the strict convexity of the value function in disadvantageous monetary payoff inequality. This "inequity-as-loss"-model thus combines elements of two

[9] There was yet one session where only 28 participants showed up and this session was with the BASE/HIGH/LOW-order of blocks. In this session we allocated 14 participants to each role, and 14 participants to each of the two outcome conditions, RISKY and CERTAIN. In that session, the BASE block lasted only for one round, and the session only lasted for 7 rounds.

[10] We also conducted two sessions with 32 student subjects in each without settlement negotiations (this data is used only in Table 2 and 3 in the online appendix). Data from these two sessions enables us to check possible spillover effects of negotiations on escalation decisions.

[11] See Brekke et al. (2016) for an experiment testing a loss-aversion model in an alternating-offer setup.

celebrated models of behavioral economics, prospect theory and inequity aversion. The key modification is to allow the utility to depend on the monetary payoff of another agent and to model the loss from disadvantageous inequity in a non-linear manner. Such ideas have earlier been put forward by Lowenstein et al. (1989), Cox et al. (2007), and Bellemare et al. (2008). Non-linearities in inequity aversion can also be found in the work of Bolton (1991) and Bolton and Ockenfels (2000).

3.1 Self-interest

Let us study the implications of the model beginning with the special case of self-interest, $\lambda = 0$. Sequential rationality (subgame perfect Nash equilibrium) suggests that the proposed and vetoed shares in the negotiation stage should depend on the expected escalation stage payoffs. The lowest offer the opponent is willing to accept makes her (almost) indifferent between accepting and vetoing it. For a risk-neutral negotiator, a share that makes the responder indifferent is equal to the expected payoff from the game ensuing to the escalation stage.[12] To secure a deal, the provoker must be offered more than her conflict payoff which equals expected return (1) or Y depending on whether it is optimal to escalate conflict or not. Similarly in a deal, the defender must be offered expected return (2) if escalation is optimal or X if it is not.

In BASE, the provoker's probability of winning is so high and the cost of escalation so low that the optimal choice calls for escalation. Her expected payoff from escalation (1) exceeds the payoff from not escalating, Y. In the negotiations stage, a self-interested sequentially rational provoker should therefore accept all offers weakly greater than her expected return from escalation (1). To the contrary, P's probability of winning in the LOW case is so low that it is suboptimal to escalate, whereas the escalation cost in HIGH is so high that it is again suboptimal to escalate. Recall that the expected return to provoker from conflict escalation (1) is equal between LOW and HIGH. Thus in HIGH and LOW, a rational provoker should accept all offers exceeding Y. Consequently, the unique subgame perfect equilibrium between risk-neutral self-interested parties predicts conflict in HIGH and LOW: a defender should never propose a positive amount or accept anything less than the full value 200 since she expects to receive 200 in case of disagreement knowing that a rational provoker never escalates. Similarly a provoker should not propose more than 190 or accept less than 10 since she will receive 10 in case of disagreement. Therefore, the value Y has the effect of slightly perturbing the balance to point out some limits of sequential rationality and subgame perfection in empirical work. This additional payoff has no impact on the optimality of escalation itself. If escalation patterns are unaffected and Y is negligibly small relative to the stakes of negotiation to much influence the negotiation patterns, then the subgame perfect equilibrium predictions of self-interested individuals may not hold. Theoretically the impact

[12] Notice that even a risk-averse opponent would accept this offer which is clearly greater than the certainty equivalent of the escalation lottery.

$r = 0.4,\ Y = 10$	BASE $p = 0.7,\ L = 10$	HIGH $p = 0.7,\ L = 58$	LOW $p = 0.1,\ L = 10$
Negotiation disagreement rate	0%	100%	100%
P's escalation choice (escalation rate)	Escalate (100%)	Do not escalate (0%)	Do not escalate (0%)
P's MAO	56	10	10
D's MAO	134	200	200
Sum of MAOs	$190 < X$	$210 > X$	$210 > X$

Table 2: Subgame perfect Nash equilibrium predictions.

is drastic, however: with the introduction of small Y, conflict becomes the only rational negotiation solution (subgame perfect equilibrium) of the game in HIGH and LOW. Conflict in HIGH LOW has a further benefit of making escalation choices to bear more impact, which is our core interest.

The subgame perfect equilibrium with self-interest predicts that cases never settle and provokers never escalate in HIGH and LOW while cases will always settle and provokers always escalate in the BASE. This may at first sight appear counter-intuitive. Yet, this is exactly what should be expected: if the provoker does not have a credible threat to escalate conflict, then a rational defender will never have to settle. On the other hand, conflict escalation is a credible threat in BASE, and a rational defender would therefore expect conflict escalation if negotiations fail. Costly conflict escalation creates room for a bargaining solution as the parties can avoid costs of escalation, which should increase the likelihood of an agreement, again something correctly captured by the subgame perfect equilibrium.

The subgame perfect Nash equilibrium predictions are summarized in Prediction 1 and in Table 2.

PREDICTION 1 *The subgame perfect equilibrium with risk-neutral self-interest predicts that*

1. *escalation is optimal in BASE,*
2. *escalation is suboptimal in HIGH and LOW ,*
3. *disagreement rate is 0% in BASE and 100% in HIGH and LOW.*

3.2 Equity reference with diminishing marginal sensitivity

Let us then consider inequity-as-loss preferences with $\lambda > 0$. Now, if the provoker does not escalate conflict, the payoff inequality is drastic and the provoker's intrinsic cost is $\lambda 190^{\gamma}$. When the provoker escalates, the payoff inequality becomes smaller if the provoker wins: the intrinsic cost is reduced to $\lambda 30^{\gamma}$. If the provoker loses, the payoff difference remains unchanged and the intrinsic cost is $\lambda 190^{\gamma}$. Escalation thus provides an opportunity to attempt to reduce payoff inequality. Due to the lower expected payoff inequality when escalating conflict, if a rational self-interested provoker prefers escalation, so does a provoker with a behavioral value function.

PREDICTION 2 *It is always optimal to escalate in BASE for all values of* $\lambda \geq 0$ *and* $\gamma \leq 1$.

Moreover, in the HIGH treatment, there is more likely to be less payoff inequality than in the LOW treatment since in the HIGH treatment the probability that the provoker wins and gets compensated is higher than in the LOW treatment. Therefore, an agent who dislikes payoff inequality is more likely to escalate in the HIGH treatment than in the LOW treatment.

PREDICTION 3 *It is optimal to escalate in HIGH and LOW when* λ *is sufficiently high and* γ *is sufficiently low, and escalation is suboptimal otherwise. The set of parameter values for which escalation is optimal in HIGH is a superset of the set of parameter values for which conflict escalation is optimal in LOW (more conflict escalation in HIGH than in LOW).*

Since utility is convex in own monetary payoff, the provoker is a risk-lover when deciding on whether to escalate conflict. She likes to gamble to reduce inequality between herself and the defender.

PREDICTION 4 *There is more escalation in RISKY than in CERTAIN.*

This prediction contrasts with what most of the conventional theoretical analysis of settlement would predict, because the typical assumption in the conventional case is that agents are risk averse or risk neutral.

The inequity-as-loss model makes yet another prediction which concerns both of the treatment variation dimensions as follows. Since the provoker's utility is convex in inequity, the escalation rate should be higher in the HIGH condition with risky escalation outcomes than in the LOW condition with certain outcomes.

PREDICTION 5 *The escalation rate in HIGH RISKY is higher than in LOW CERTAIN.*

Notice the difference between Prediction 5 and Prediction 3. The latter holds that escalation rate is higher in the HIGH condition than in the LOW condition ceteris paribus. Prediction 5 states that there is a difference in escalation rates between HIGH RISKY and LOW CERTAIN.

The above predictions are the key theoretical predictions, concerning the escalation rates, stemming from the behavioral value function. Derivations of these are in the online appendix.

We also derive predictions regarding the disagreement rates assuming that the preference parameters are complete information (the proof is provided in the online appendix).

PREDICTION 6 *The subgame perfect equilibrium with inequity-as-loss model (with complete information) makes the following disagreement rate predictions:*

- *in BASE, the disagreement rate is 0% independently of* λ *and* γ,

- *the disagreement rate is lowest in BASE and highest in LOW,*
- *in HIGH, the disagreement rate is lower in RISKY than in CERTAIN.*

In a model version where preference parameters are incomplete information, the proposer would trade off a larger share to herself against a lower probability of agreement when deliberating which proposal to make. In such a model, disagreement rates would typically never equal 0% or 100%, but it would still predict the extreme 100% escalation rate in BASE. Stochastic choice is an alternative to incomplete information in introducing a trade-off between proposer's own share and the probability of acceptance and thus in reaching less extreme disagreement rate predictions, and even in BASE. The logit quantal response equilibrium for instance assumes logistically distributed choice probabilities and that the variation in the choices is correctly predicted by other players. We investigate this in Section 4.3.

4 Results

In this section we scrutinize the empircial observations and how they relate to the theoretical predictions. The numbering of the results broadly matches with the associated predictions in the previous section.[13] We use variably the following methods to examine the robustness of our results: non-parametric tests (mainly test of proportions and Mann-Whitney U-tests, sometimes χ^2 tests and even within-subject McNemar or Wilcoxon signed rank tests), random-effects linear probability regressions clustering standard errors at the individual level (GLS), subsession level and sometimes even at the session level. In the linear probability regressions, we do further robustness analysis with "wild boostrapped" standard errors (Cameron et al. 2008), and linear probability regressions treating period as a linear trend instead of period-dummies, and with random effects logistic regressions with the same clustering structure. We report the linear regressions with period dummies in the main tables and indicate non-robustness with respect to the alternative specifications in the main text and footnotes when present.[14] We perform a comprehensive set of statistical tests for each hypothesis to provide the reader a good understanding of the robustness of our results with respect to the statistical specifications and with respect to experience.

Regarding the non-parametric tests, notice that in the first round, each P's escalation choice constitutes an independent observation, or alternatively in the analysis of disagreement, each proposal or MAO of a player in a given role or each agreement-disagreement outcome of each P-D pair constitutes an independent observation.

[13] In the online appendix, we report the escalation rates of some additional treatments where the whole experiment consisted of the provoker's escalation choice. There was only a passive defender-recipient but no negotiations. These additional treatments are excluded from the main analysis.

[14] Angrist and Pischke (2009, sections 3.3 and 3.4) argue against non-linear alternatives and in favor of the linear specification when a saturated model with randomized treatments is used in a panel data setting. The linear regression coefficients give in a straightforward manner the differences in the treatment averages controlling for period effects.

In later rounds, the feedback given in the first round about the behavior of the randomly matched participant in the opposing role potentially influences behavior within the subsession in the later rounds. The perfect strangers matching mitigates the effect. Nevertheless, the average escalation rate over the Ps of a subsession, the average proposal or MAO of players in a given role of a subsession, or the disagreement rate over the P-D pairs of a subsession constitutes an independent observation in the last round. Each observation is a binary variable in the first round and a percentage in the last block. Therefore, we use non-parametric Test of proportions and Mann-Whitney U-test with the first round escalation choices and the average escalation rates within the last block and the significance refers to these, respectively, if not mentioned otherwise. The tests are one-sided whenever the theoretical prediction gives a reason to expect a directional effect. We test whether the observed escalation behavior is significantly different from the predicted one, and whether the behavior is significantly different across treatment conditions.

Since theoretical rational equilibrium notions are typically thought to characterize steady-state behavior once learning has taken place, it is of interest to check whether there are differences in escalation behavior over time.[15] This is done by means of comparing the non-parametric tests with the first period and last block data. Notice that we have only one eighth of the independent observations in the tests in the last block than in the tests in the first round, but the variable in the last round is continuous rather than binary and the tests are different in nature. Therefore, there are differences in the statistical power of the first round and last block tests. The significance differences between first period and last block tests may reflect these differences in addition to the learning effects.

4.1 Provoker's conflict escalation choices

In this section, we analyze provokers' conflict escalation decisions. The observed escalation rates across the various treatment conditions are given in Table 3, and there are three main observations. First, there is more conflict escalation in BASE than in HIGH and LOW where the parameters are less propitious to conflict escalation (on the bottom line of Table 3, compare the second with the third and the fourth column). Second, there is also more conflict escalation in RISKY than in CERTAIN (in the last column of Table 3, compare the cells on line two and three). Third, the escalation rate is higher in HIGH RISKY than in LOW CERTAIN (compare the cell in column three and line two with the cell in column four and line three).

Let us begin with Prediction 1**.** That the observed escalation rate is higher in BASE than in HIGH or LOW in Table 3 is in line with the comparative statics predictions of self-interested rationality. Yet, our main treatments data exhibits an abundance of choices not maximizing expected monetary return. In BASE, 14% of the provokers do not escalate conflict although they should (the cell in the second column of the last row

[15] See for instance Friedman (1953, pp. 192-193) and Thaler (1980, pp. 57-58).

$r = 0.4,\ Y = 10$	BASE, $p = 0.7,\ L = 10$	HIGH, $p = 0.7,\ L = 58$	LOW, $p = 0.1,\ L = 10$	Total
Risky escalation outcomes	89%	73%	64%	73%
Certain escalation outcomes	83%	48%	48%	56%
Total	86%	60%	56%	65%

Table 3: Empirical escalation rates pooled over all rounds

in Table 3; the subgame perfect Nash equilibrium prediction is that 100% escalate). With prohibitively high costs (HIGH), still 60% of the provokers escalate (the subgame perfect Nash equilibrium prediction is 0%), and similarly 56% of the provokers do so with prohibitively low probability of winning, LOW (the subgame perfect Nash equilibrium prediction is 0%). Figure 1 in the online appendix illustrates how escalation rates evolve over time, conditional on treatment conditions.

The escalation rate in BASE is not significantly different from 100% with first round data but it is significantly different at the 1% level when using subsession averages in the last block. The escalation rates in the HIGH and LOW conditions are significantly different from 0% at 1% level both with first round and with last block data. Thus our data rejects Prediction 2 and part 1 of Prediction 1 with experienced participants and part 2 of Prediction 1 with both inexperienced and experienced participants.

RESULT 1 & 2 *The escalation rate of both experienced and inexperienced participants is significantly different from 0% in HIGH and LOW. The escalation rate of experienced participants is significantly different from 100% in BASE.*

Escalation rates being significantly above 0% is consistent with Prediction 3 which states that escalation is optimal for a range of parameter values λ and γ. Prediction 2 implies that escalation rate should be 100% in BASE (i.e. same prediction as part 1 of Prediction 1), whereas Prediction 3 implies that escalation rate should be between 0% and 100% in HIGH and LOW so that escalation rate in HIGH should be weakly higher than in LOW.

To scrutinize Predictions 2 and 3 further, we report the results of a regression analysis in Table 4. A dummy variable indicating whether the provoker chose to escalate (1) or not (0) is regressed on treatment variables HIGH, LOW, and RISKY and their interactions controlling for the period of negotiations. The baseline in this regression is our BASE condition with certain escalation outcomes. The standard errors are clustered at subsession level in the main specification, and these are reported in brackets on the second line below the regression coefficients. The significance of coefficients at 1%, 5% and 10% level are reported in the usual manner with respect to this clustering specification. However, we also provide the standard errors clustered at the individual level in parentheses right below each regression coefficient. We also ran the

regressions with session level clustering, but the results were robust to the clustering specifications.

The table shows that the escalation rate in BASE is significantly higher than in HIGH. Similarly, the escalation rate in BASE is significantly higher than in LOW. This is indicated in Table 4 by the negative and significant (at the 1% level) HIGH and LOW coefficients in all of our regression models whether we regress escalation merely against the plain treatment effects and period dummies (models 1) or whether we allow for interactions between the treatment variables (models 2) or even allow the treatments to interact with the period dummies (models 3 and 4).[16] The significance of the results remains the same independently of the level of clustering. We also ran the linear probability model regressions in Table 4 with wild boostrapped standard errors, and separately with linear time trends instead of treating period as a dummy, and even letting the trend interact with treatments in the same manner as in models 3 and 4. All the significant coefficients remain significant and their magnitudes are unchanged in these alternative specifications. These comparative statics of the treatments effects are in line with Predictions 1 and 2.[17]

Corresponding one-sided χ^2 tests yield p-values of 0.003 and 0.0002 when comparing BASE to HIGH and BASE to LOW, respectively (first-round data only where each provoker is an independent observation). The escalation rate in BASE is also significantly higher than in HIGH (p-value 0.034) and significantly higher than in LOW (p-value 0.0002) when comparing subsession averages in the last block. We also carried out within-subsession tests of escalation rates comparing BASE vs HIGH, on the one hand, and BASE vs LOW, on the other. For this purpose for each subsession, we classified the average escalation choice of the provokers at the last round of each block according to the corresponding condition (one average escalation choice from BASE, HIGH and LOW conditions, respectively, for the provokers of each subsession) and ran one-sided Wilcoxon signed rank test comparing two of the three averages at a time within a subsession. We find that escalation rate is significantly higher in BASE than in HIGH (p-value 0.002), and that escalation rate is statistically significantly higher in BASE than in LOW (p-value, 0.002), too. In conclusion, our results provide support for the comparative statics derived from Predictions 2 and 3.

RESULT 3 *Escalation rate is higher in the BASE treatment than in the HIGH and LOW treatments.*

Prediction 4 states that escalation rate in the RISKY condition should be higher than in the CERTAIN condition. The top two rows of Table 3 give an indication that there is an effect to the predicted direction. To explore Prediction 4 more carefully, we first compare the first-round escalation rates in RISKY and CER-

[16] In Figure 1 in the online appendix, we illustrate how the behavior evolves over periods and treatment blocks. There are some differences in the trends across blocks and thus models 3 to 4 provide robustness checks for the treatment effects.

[17] Results of these additional regressions available upon request. We also ran thre regressions with non-linear logit-specifications and calculated the average marginal effects for that specification. These are for HIGH = -0.276, and for LOW = -0.317. The marginal effect of RISKY equals 0.166. These are remarkably close to the effect estimates of the linear model (1) in the Table 4. Yet, it is worth noting that this laboratory study focuses on the significance of the qualitative treatment effects, rather than the absolute effect sizes.

Escalation	(1)	(2)	(3)	(4)
RISKY	0.170***	0.053	0.059	0.053
	(0.045)	(0.047)	(0.076)	(0.046)
	[0.050]	[0.067]	[0.080]	[0.062]
HIGH	-0.245***	-0.343***	-0.354***	-0.369***
	(0.036)	(0.055)	(0.054)	(0.092)
	[0.040]	[0.062]	[0.056]	[0.054]
RISKY × HIGH		0.196***	0.219***	0.188***
		(0.076)	(0.071)	(0.072)
		[0.071]	[0.066]	[0.070]
LOW	-0.292***	-0.347***	-0.359***	-0.441***
	(0.039)	(0.054)	(0.054)	(0.088)
	[0.038]	[0.047]	[0.041]	[0.057]
RISKY × LOW		0.112	0.135**	0.113
		(0.076)	(0.078)	(0.077)
		[0.074]	[0.068]	[0.070]
Constant	0.766***	0.824***	0.820***	0.874***
	(0.043)	(0.046)	(0.056)	(0.062)
	[0.049]	[0.060]	[0.059]	[0.041]
Period dummies	YES	YES	YES	YES
RISKY × Period dummies	NO	NO	YES	NO
LOW × Period dummies	NO	NO	NO	YES
HIGH × Period dummies	NO	NO	NO	YES
Observations	1,250	1,250	1,250	1,250

Robust standard errors clustered by individual subjects in parenthesis

...and by subsession in brackets.

*** p<0.01, ** p<0.05, * p<0.1. Significance level is based on subsession clustering.

Table 4: Escalation rate, linear probability regressions, main treatment effects.

TAIN (see Figure 1 in the online appendix for an illustration of escalation rate differences across treatment conditions, blocks and periods). The escalation rate in RISKY is significantly higher than in CERTAIN. The one-sided χ^2 test gives a p-value of 0.004. These tests exploit the first-round data only. To understand whether the pattern is robust to experience, we took subsession averages of the provokers' escalation choices and compared the subsession average escalation rates in the last block between RISKY and CERTAIN. This difference is significant at 5% level (p=0.047) which suggests that the effect is robust to experience.

The regression analysis in Table 4 provides another approach for examining Prediction 4. This analysis provides evidence that there is more conflict escalation in RISKY than in CERTAIN: in the non-interacted model (first column) the coefficient of variable RISKY in Table 4 is positive and significant.[18] In the additional regressions (models 2 to 4), we allow for interactions between HIGH or LOW and the riskiness of escalation outcomes, RISKY (model 2), and even interactions of the treatments with period (models 3 and 4).[19] We find that the the coefficient of RISKY is always positive but statistically significant only in the first regression

[18]The p-value should be adjusted for multiple hypothesis testing (List et al., 2016). Result 1 would be statistically significant at 5% level for maximally six simultanous tests (for the null hypothesis that none of them are significant, assuming independence). Thus the result can be seen robust to multiple simultaneous testing.

[19]Such interactions are conceivable since we have within-subject variation across blocks.

model. The interacted models reveal that it is the HIGH condition with risky escalation outcomes that is particularly prone to escalation (positive and highly significant coefficient of RISKY x HIGH in models 2 to 4).[20] The significance of the RISKY coefficient in the non-interacted model seems to be mainly driven by the effect of RISKY in the HIGH condition; the coefficient of RISKY is insignificant in models 2 to 4 where the interactions are allowed. In conclusion our results provide support for Prediction 4.

RESULT 4 *Escalation rate is higher in RISKY than in CERTAIN.*

Prediction 3 holds that there should be more escalation in HIGH than in LOW, and Prediction 5 states more specifically that the escalation rate is higher in HIGH RISKY than in LOW CERTAIN. Before turning to formal tests, let us discuss our evidence so far from the perspective of Prediction 3 and Prediction 5. Regarding Prediction 3, the escalation rate is higher in HIGH than in LOW when escalation outcomes are risky in Table 3 (compare the cells in column three and four on line two), but not when outcomes are certain (compare the cells in column three and four on line three). In Table 4, the coefficient of LOW is more negative than that of HIGH in regression model 1 with plain treatment effects without interactions. This is again indicating that escalation rates are higher in HIGH than in LOW. Moreover non-parametric tests of whether escalation rate in BASE is higher than in HIGH or LOW are more significant when comparing BASE with LOW than when comparing BASE with HIGH (both with first round data and data from the last block between subjects or subsessions, and with within-subsession tests, see above before Prediction 3). Regarding Prediction 5, the escalation rate is higher in HIGH RISKY than LOW CERTAIN in Table 3 (compare the cell in column three and line two to the cell in column four and line three) and in Table 4 the coefficient of the interaction term in model 4, RISKY $\times$ HIGH, is significant and positive but the interaction term RISKY $\times$ LOW is not.[21] Thus there is some evidence in favor of the predictions. None of these are however clean tests of them.

To test Prediction 3 further, we first carry out non-parametric tests. In addition to the between-subjects test of proportions with the first period escalation choices and the between subsessions Mann-Whitney U-test with the escalation rate in the last block of a subsession, we also ran a within-subsession one-sided Wilcoxon signed rank test comparing the average escalation rates among the provokers of a subsession across the last rounds of each block. When comparing experienced participants in the last block, there is a significantly higher escalation rate in HIGH than in LOW when outcomes are RISKY ($p=0.023$), but not when they are CERTAIN. None of the other tests are significant.

[20] In models (2) and (4), the significance is not robust to the logit specification, but nevertheless, probit gives significant coefficients.

[21] In model 3 of Table 4, the coefficient is significant at a 5% level in model 3 but not in other models. In Logit versions of regressions 2 to 4, the coefficient RISKY $\times$ LOW is never significant.

We also ran linear probability model regressions using the data from HIGH and LOW only, and excluding the data from the BASE. In Tables 2 and 3 in the online appendix, the reference condition is LOW with CERTAIN and RISKY escalation outcomes, respectively and the covariates are the treatment effects. In the regressions in the first two columns, we do not allow for interactions. Regarding prediction 4 in Table 2 in the online appendix, the coefficients CERTAIN and RISKY are significant and in line with Prediction 4. The variable HIGH in Tables 2 and 3 in the online appendix capture the difference in the escalation rates between HIGH and LOW in RISKY and CERTAIN, respectively. The effect is significant in neither case and thus gives no support to Prediction 3. Thus our only statistically significant support for Prediction 3 comes from the within-subject non-parametric tests with the main data, and from the regression analysis where additional data from sessions without negotiations was included.[22]

RESULT 3' *Escalation rate is not significantly higher in HIGH than in LOW.*

To study Prediction 5, we ran the joint test that the coefficients HIGH, RISKY and the interaction term RISKY $\times$ HIGH are jointly all zero in the linear probability model (GLS) in the second column of Table 3 in the online appendix. We reject the null hypothesis at 1% level (p-value = 0.0005). The result is robust to the alternative specifications of the linear model (period dummies vs. trends, clustering at the individual or session level). We also run non-parametric tests to study whether escalation rate is higher in RISKY than in CERTAIN in HIGH and LOW separately. In HIGH, the effect of RISKY is significant both with first period data (p=0.018, one-sided) and with data from the last block (p=0.023, one-sided). In LOW, the effect of RISKY is significant at 10% level with first period data (p=0.067) and not significant with data from the last block (p=0.2). In summary, the non-parametric tests suggest a weakly positive but insignificant effect of HIGH (Result 3'), and a positive and significant effect of RISKY (Result 4) which is somewhat stronger in HIGH than in LOW. Thus our evidence supports Prediction 5.[23]

RESULT 5 *Escalation rate is significantly higher in HIGH RISKY than in LOW CERTAIN.*

4.2 Negotiations and disagreement

Let us then turn our attention to the disagreement rates. Part 3 of Prediction 1 and Prediction 6 summarize the predicted treatment effects on disagreement rates based on the subgame perfect equilibrium of the self-interested players and inequity-as-loss-motivated players respectively.

[22] Including additional data to the regression from a session where no negotiations took place before the escalation decision, we do find a significant effect of HIGH in RISKY and no effect in CERTAIN, thus giving partial support to Prediction 3.

[23] Including additional data to the regressions from a session where no negotiations took place before the escalation decision, provides further support to Prediction 5.

$r = 0.4,\ Y = 10$	BASE $p = 0.7,\ L = 10$	HIGH $p = 0.7,\ L = 58$	LOW $p = 0.1,\ L = 10$	Total
Risky escalation outcomes	51%	38%	47%	45%
Certain escalation outcomes	55%	37%	50%	46%
Total	53%	37%	49%	45%

Table 5: Disagreement rates, pooled over all rounds

Table 5 reports the disagreement rates in our six different experimental conditions. The disagreement rate is lower in the HIGH treatment than in the other two, but there are no major differences in the disagreement rates between RISKY and CERTAIN (compare cells on the second against those on the third row in Table 5). Histograms in Figures 2 and 3 capture the frequencies of the provokers' and the defenders' offers, respectively, for the three conditions in CERTAIN (three bottom panels) and the three conditions in RISKY (top panels) over all rounds.[24] The upwards-sloping line in each subgraph depicts the empirical cumulative distribution of MAOs, i.e. the aggregated acceptance probability in the population of agents on the opposing side. Notice that the offer of 100 that shares the pie equally if accepted tends to be the modal offer, but the offers to the defenders (Figure 2) are higher than the offers to the provokers (Figure 3) reflecting the higher conflict payoffs that the defenders receive in all conditions whether or not the provoker escalates the conflict. In fact the modal offer to the provokers is 80 rather than 100 in many conditions and there is also much more dispersion in the offers to the provokers. The majority of MAOs are set between 80 and 100. The defenders set higher MAOs and make lower offers than the provokers, again in line with their higher conflict payoffs and thus with the comparative statics prediction of self-interested sequential rationality.[25]

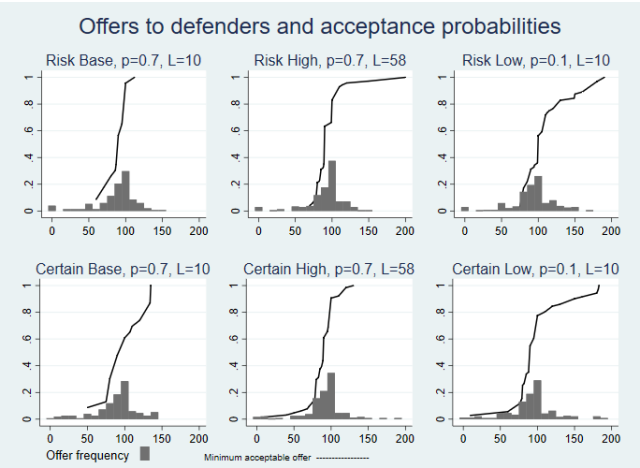

Figure 2: Offers to defenders, all rounds

[24] In Figure 1 in the online appendix, we illustrate the evolution of negotiation and escalation behavior and disagreement rates over time, conditional on treatment conditions.

[25] Between-subjects Mann-Whitney U-test with both first period data and data with last block confirm that the defenders' MAOs (offers) are significantly higher (lower) than those of the provokers at (1% level).

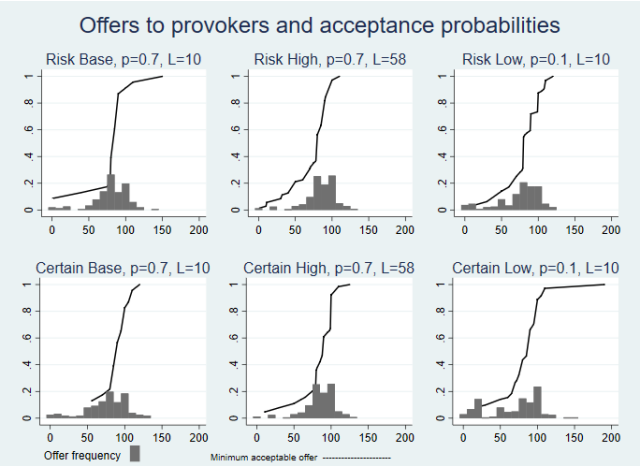

Figure 3: Offers to provokers, all rounds

A key observation from Figure 3 is that the distributions of offers to the provokers seem to have a fatter left-tail in BASE and LOW than in HIGH where the left-tail is almost absent. The defenders seem to have adopted a less aggressive bargaining strategy in HIGH. This is likely to contribute to the lower disagreement rate in HIGH. Similarly in Figure 2, the empirical cumulative distribution of the MAOs of the defenders rises steeply at the 100-threshold in the two middle panes of HIGH; the curve is yet flatter at the equally-splitting 100 in LOW conditions (two right panes) and in the certain BASE condition (bottom left pane). This is indicating that defenders are more likely to reject offers above the 100-100 split in LOW and BASE than in HIGH, again an indication of a less aggressive strategy in HIGH.

There are no significant differences in the defenders' proposals when comparing between HIGH and BASE or HIGH and LOW using the first round data (each individual proposal as an independent observation). Yet, using the average proposals of the defenders in each last block of a subsessions as an independent observation and comparing across conditions, we find that the defenders' offers to provokers are significantly smaller in LOW than in HIGH (at the 1% level). No significant difference is observed between BASE and HIGH, neither in the first round nor in the last block. The MAOs of the defenders are significantly lower in HIGH than in LOW using both first round data (1% level) and last block data (1% level). There is no significant difference in the first period between HIGH and BASE, but in the last block the average MAOs of the defenders are significantly lower in HIGH (1% level). Thus the defenders are significantly less aggressive in HIGH than in the other two, and moreover this tendency becomes stronger with experience.

There are no significant differences neither in the MAOs nor in the proposals of the provokers across treatments, not in the first round nor in the last block, which again conflicts the subgame perfect equilibrium predictions. We will return to these observations in the next subsection.

To study more rigorously how the differences in negotiation strategies across treatments are reflected in negotiation outcomes, we run a panel regression analysis. Since agreement or disagreement is an outcome of a bilateral negotiation, the number of observations in these regressions is equal to the numbers in the regressions in Section 4.1 where the escalation behavior of the provokers in each pair are analysed. In Table 6, a dummy variable indicating disagreement is regressed on treatment variables controlling for the period of play and, in some models, allowing even for interactions between treatments and period. The baseline in this regression is our HIGH condition with certain escalation outcomes which has the lowest disagreement rate. The standard errors are clustered at the subsession level and they are reported in parenthesis and the significance of the regression coefficients are given in regard to this clustering specification. The standard errors clustered at session level are also given in brackets.

The regressions confirm that the disagreement rate is significantly lower in HIGH than in the other two. All regressions have positive and significant coefficients for both BASE and LOW (BASE and LOW variables) indicating higher disagreement rates compared to the HIGH.[26] The third and fourth regression looks more carefully at the effects of risky escalation outcomes by adding interaction effects between BASE and LOW with RISKY. None of the interaction effects are significant. This evidence is consistent with our findings above that the defenders are less aggressive in HIGH than in the other two, especially in the later rounds.

Since the escalation rates are higher in the risky conditions, one could expect that disagreement rates might be lower in those conditions than in the certain ones - whenever the escalation rates are higher, there is an incentive to attempt avoiding the escalation costs by agreeing. Yet, the coefficient RISKY in Table 6 is not significant although the coefficient is negative in the non-interacted models in the first two columns. All non-parametric tests comparing the disagreement rate in RISKY and CERTAIN conclude with insignificant effects.

It is of interest to relate the patterns of Table 5 to the part 3 of Prediction 1 and Prediction 6 where the predictions of the subgame perfect equilibrium in the self-interest and inequity-as-loss model were derived, respectively. The predictions miss the mark on a number of dimensions: First, the disagreement rate in BASE is 53%, well above the 0% predicted by both Prediction 1 and Prediction 6. Likewise, the disagreement rates in HIGH and LOW are well below the 100% predicted by Prediction 1.[27] Second, the disagreement rate is not lowest in BASE, as predicted by Prediction 1 and 5, but in HIGH condition.[28] Third, the disagreement rate is not lower in BASE than in HIGH, as predicted by Predictions 1 and 6.[29]

[26] All results in Table 6 are robust to the the Logit specification. We also calculated the average marginal effects for the Logit-version of the model in the first column. These are for BASE = 0.1604, for LOW = 0.1117, and for RISKY = -0.0144.

[27] In BASE, the disagreement rate is statistically significantly different from zero at 1% level both with first round data and with last block data. In both HIGH and LOW, the disagreement rate is statistically significantly different from 100 at 1% level both with first period data and with last block data.

[28] The disagreement rate is also not highest in LOW but in BASE. Yet, there is only a 4 percentage point difference in the disagreement rates, so they are virtually equal.

[29] The regression results discussed above show that the difference between BASE and HIGH is statistically significant (the

RESULT 6 *The subgame-perfect equilibrium with both self-interest and with inequity-as-loss are inconsistent with the observed disagreement rates.*

The next section shows that the QRE version of the inequity-as-loss model makes more accurate predictions.

Disagreement	(1)	(2)	(3)	(4)
RISKY	-0.014	0.012	0.157*	0.012
	(0.043)	(0.052)	(0.090)	(0.052)
	[0.036]	[0.039]	[0.089]	[0.035]
BASE	0.148***	0.172***	0.160**	0.037
	(0.043)	(0.064)	(0.067)	(0.094)
	[0.047]	[0.063]	[0.069]	[0.087]
RISKY×BASE		-0.052	-0.019	-0.052
		(0.080)	(0.084)	(0.076)
		[0.070]	[0.079]	[0.067]
LOW	0.112***	0.130***	0.135***	0.079
	(0.034)	(0.057)	(0.051)	(0.115)
	[0.037]	[0.060]	[0.052]	[0.079]
RISKY×LOW		-0.038	-0.036	-0.032
		(0.068)	(0.057)	(0.067)
		[0.065]	[0.048]	[0.065]
Constant	-0.383**	-0.370***	0.295***	0.416***
	(0.055)	(0.058)	(0.077)	(0.067)
	[0.049]	[0.051]	[0.079]	[0.078]
Period dummies	YES	YES	YES	YES
RISKY × Period dummies	NO	NO	YES	NO
LOW × Period dummies	NO	NO	NO	YES
HIGH × Period dummies	NO	NO	NO	YES
Observations	1,250	1,250	1,250	1,250

Robust standard errors clustered by subsession in parenthesis

Robust standard errors clustered by session in brackets

*** p<0.01, ** p<0.05, * p<0.1 (based on subsession clustering)

Table 6: Disagreement rate regressions.

4.3 Explanatory power of the logit quantal response equilibrium model

The behavioral predictions of Section 3.2 predict the comparative statics between the conditions better than the self-interest model of Section 3.1. This holds true for the escalation rate predictions in particular. Yet, when it comes to the disagreement rate predictions, both the model with self-interest and the behavioral model fail to predict correctly the comparative statics between HIGH and BASE.

coefficient BASE is positive and significant, rather than negative and significant, in the regressions). There is no significant difference in the disagreement rates between BASE and HIGH in the first period, but in the last block the disagreement rate in BASE is significantly higher than in HIGH at 10% level (p=0.051). Thus the evidence is rather opposite to the prediction. The tests between HIGH and LOW and between BASE and LOW are not significant either with first round or with last block data. The shortcoming with these tests is that, in the first period, the parties (the defenders in particular) have not yet had an opportunity to learn about the implications of disagreement, and in the last block, there is only a limited number of independent observations. As shown above, the defenders' bargaining behavior seems to be considerably less aggressive in the last rounds.

While providing a useful benchmarking role for understanding behavior, the complete information subgame perfect Nash equilibrium turns out too precise and extreme for providing the best fit with data, because for given preference parameters the model always predicts a 0% or 100% escalation rate and a 0% or 100% disagreement rate. There are alternatives to the Subgame-perfect equilibrium which predict variance in escalation and agreement patterns. One can introduce incomplete information (Bayesian Nash equilibrium) about other-regarding preference, stochastic choice (logit quantal response equilibrium), or unobservable heterogeneity in preferences (random utility). All approaches rely on strong distributional assumptions: either one must make strong assumptions about the commonly known preference parameter distributions (Bayesian Nash equilibrium), or about how choice probabilities are related to the expected payoffs of the game (logit QRE), or about the nature of preference heterogeneity not observable to the researcher (random utility).

Wheareas all approaches are conceivable, we will show next that nearly all of the qualitative comparative statics patterns of the key treatment conditions can be fairly well accommodated using the notion of (agent) logit quantal response equilibrium (McKelvey and Palfrey, 1998). In the logit-QRE the choice probabilities reflect rationality in the sense that they are inversely related to the opportunity costs of the choices and the implied choice probabilities are correctly anticipated by the opponents. This relatively small departure from perfect rationality allows us to drastically improve the settlement and escalation rate predictions. This general idea has proved successful in a number of other strategic interaction situations (see Goeree and Holt, 2001; Goeree et al. 2016) but our setup studying the risks and costs of escalation at negotiation impasse is novel. In addition to the comparative statics, we provide the maximum likelihood estimates for the parameters of the model and the implied escalation and disagreement rates.[30]

In the logit quantal response model, the choice probabilities are proportional to the exponentials of the expected utilities, v_i, of the actions given the beliefs on the opponents' behavior. Let us denote the expectation of i about the action profile a_{-i} of other players by $\hat{\sigma}^i_{-i}(a_{-i})$. In the quantal response equilibrium, player i chooses action a_i with probability

$$\sigma_i(a_i) = \frac{exp\left((1/\mu)\left(\sum_{a_{-i}} \hat{\sigma}^i_{-i}(a_{-i}) v_i(a_i, a_{-i})\right)\right)}{\sum_a exp\left((1/\mu)\left(\sum_{a_{-i}} \hat{\sigma}^i_{-i}(a_{-i}) v_i(a, a_{-i})\right)\right)}. \tag{4}$$

This formulation allows for considering both stochastic decision making by self-interested agents (replace v_i with π_i) and stochastic behavioral agents (use the value function v_i given in (3)).

[30] The comparative statics predictions of the escalation rates between the six various treatment conditions are correct (see Predictions 7 and 8 below) apart from the fact that empirically observed escalation rates in HIGH CERTAIN and LOW CERTAIN are equal (the behavioral QRE prediction holds that the escalation rate should be higher in HIGH) and the fact that disagreement rates in RISKY and CERTAIN are essentially equal (in the end the maximum likelihood model parameters in fact predict no difference in this respect).

Taking the ratio of choice probabilities of two different actions a_i' and a_i'' yields

$$\frac{\sigma_i(a_i')}{\sigma_i(a_i'')} = \frac{exp\left((1/\mu)\left(\sum_{a_{-i}} \hat{\sigma}^i_{-i}(a_{-i})v_i(a_i', a_{-i})\right)\right)}{exp\left((1/\mu)\left(\sum_{a_{-i}} \hat{\sigma}^i_{-i}(a_{-i})v_i(a_i'', a_{-i})\right)\right)}, \tag{5}$$

and thus the ratio of choice probabilities is proportional to the ratio of exponentials of expected utilities. Expectations and choice probabilities must coincide in equilibrium and thus $\hat{\sigma}^j_i = \sigma_i$ for $j \neq i$. The noise parameter μ indicates the level of error in decision makers' choices so that the smaller μ is the more responsive are the decision makers to differences in utility. As μ tends to zero, the choice probabilities converge to a Nash equilibrium of the game (with respect to the social utility specified in the motivation function v_i).

For our settlement negotiation game, it is crucial to note that under self-interest when conflict escalation is suboptimal due to high costs (HIGH), the opportunity cost of escalation is 2 for the provoker while it is 114 for the defender. Equations (4) and (5) imply that letting μ tend towards zero and thus making parties more rational in their choices, defenders tends to shy away from suboptimal negotiation strategies much faster than provoker abandons conflict escalation. When provokers tremble in their escalation decisions, the noise has a more drastic impact on the defenders' incentives in the negotiation table than it has on the provokers' incentives in HIGH: the defender's expected conflict payoff falls from 200 to 86 when the provoker shifts from no escalation to escalation, but the provoker's expected payoff falls only from 10 to 8 (or might even rise if behavioral motivation is introduced and is sufficiently important). In contrast in LOW, defenders' expected conflict payoff is 200 when the provoker does not escalate conflict and 182 if the provoker does, and therefore the logit-QRE predicts tougher defender bargaining behavior in LOW compared to HIGH. This is indeed what is observed in the data, as the disagreement rate in HIGH is lower than in the other two conditions. Defenders' softer negotiation behavior is also seen from their mean MAOs. The mean defender MAO in HIGH is 92 which is lower than in LOW (104) and BASE (102). The same is visible in their proposals to the provokers. Mean defender proposal in HIGH is 84, compared to 71 in LOW and 77 in BASE conditions.

Notice that that the above is consistent with our observations about negotiation strategies in section 4.2. It is the defenders who set lower MAOs and higher offers in HIGH condition. In addition, the tendency is stronger in later rounds as suggested by the interpretation of the QRE as a stochastic fixed point (or stationary distribution) of behavior.

Let us now continue to assume that choice probabilities satisfy equation (4) and moreover allow the provoker to have a behavioral value function at the escalation stage. In this case, we can show that in any QRE, (i) the provoker is more likely to escalate when the escalation outcomes are RISKY rather than CERTAIN, (ii) the provoker is more likely to escalate in BASE than in HIGH, and also more likely to escalate

in HIGH than in LOW. Finally we can also show (iii) that the difference in predicted escalation rates between HIGH and LOW is greater under RISKY than under CERTAIN. These are essentially the QRE-versions of the Predictions 2 to 5 above and they are consistent with data, apart from the fact that the evidence for higher escalation rate in HIGH than in LOW is very weak. The proof of the results and a formal mathematical statement of the results is provided in the online appendix.

PREDICTION 7 *In any logit quantal response equilibrium,*

1. *the provoker is more likely to escalate in RISKY than in CERTAIN,*
2. *the provoker is more likely to escalate in BASE than in HIGH, and also more likely to escalate in HIGH than in LOW.*
3. *among HIGH and LOW, the escalation rates are highest in HIGH RISKY and lowest in LOW CERTAIN.*

In addition we can show that the observed disagreement rates, which deviate strongly from the predictions of the subgame perfect equilibrium (with self-interest or behavioral motivation - see Predictions 1 and 6), are consistent with those predicted by the behavioral logit-QRE for suitably chosen parameter values.

PREDICTION 8 *In a logit quantal response equilibrium with sufficiently high λ, sufficiently small γ, and sufficiently large μ,*

1. *the disagreement rate is higher when outcomes are CERTAIN than RISKY,*
2. *the disagreement rate is higher in LOW than in HIGH,*
3. *the disagreement rate is higher in BASE than in HIGH.*

We employ maximum likelihood estimation to yield an estimate for μ (the rationality parameter), for λ (the loss aversion parameter), and for γ (the reflection effect parameter) (See Section 3). We first classify offers and MAOs as follows. Offers into 5 coarse classes rounding offers 0-49 to 0 (MAOs 1-50 to 50), offers 50-99 to 50 (MAOs 51-100 to 100) and so forth so that offers labeled as k are all compatible with MAOs labelled as k and use this coarsened empirical distribution of offers and responses to calculate the log-likelihood of the profile of choices given the values of the parameters μ, λ, and γ.[31] The corresponding choice probabilities predicted by the model constitute the unique solution of the system of equations (5).

[31] See Costa-Gomes and Zauner (2001) The coarsening is needed to facilitate the numerical calculation of the equilibrium choice probabilities and their estimation.

$r = 0.4,\ Y = 10$	BASE $p = 0.7,\ L = 10$	HIGH $p = 0.7,\ L = 58$	LOW $p = 0.1,\ L = 10$
Logit-QRE, paramet. values $\mu = 55$, $\lambda = 0$	71%	49%	49%
... with $\mu = 1.2$,$\lambda = 0.45$, $\gamma = 0.95$ (risky)*	91%	76%	53%
... with $\mu = 1.2$,$\lambda = 0.45$, $\gamma = 0.95$ (certain)*	87%	68%	51%
Subgame perfect Nash with $\lambda = 0$	100%	0%	0%

Table 7: Predicted escalation rates.

$r = 0.4,\ Y = 10$	BASE $p = 0.7,\ L = 10$	HIGH $p = 0.7,\ L = 58$	LOW $p = 0.1,\ L = 10$
Logit-QRE with $\mu = 55$, $\lambda = 0$	62%	55%	61%
... with $\mu = 1.2$,$\lambda = 0.45$, $\gamma = 0.95$ (risky)*	63%	49%	68%
... with $\mu = 1.2$,$\lambda = 0.45$, $\gamma = 0.95$ (certain)*	63%	52%	68%
Subgame perfect Nash with $\lambda = 0$	0%	100%	100%

Table 8: Predicted disagreement rates

When constraining to self-interest (imposing $\lambda = 0$), the maximum-likelihood μ estimate thus received is $\mu^* \approx 55$ which gives the best fit with the data. The corresponding disagreement rates and the associated empirical frequencies are given in Table 5.[32] The logit-QRE with $\mu = 55$ asserts that disagreement rate should be lower in HIGH than in the other two. This prediction is borne out by data. Remarkably, the logit-QRE also predicts correctly that the disagreement rate is approximately equal in LOW and BASE.

We can further improve maximum likelihood and moreover capture the comparative statics of the disagreement rates, and particularly of the escalation choices, by introducing behavioral motivations as characterized by equation (3). In this case the maximum likelihood parameter estimates yield $\mu = 1.2$, $\lambda = 0.45$, and $\gamma = 0.95$. These parameters are consistent with the proposed other-regarding loss aversion model with a value function which is convex in π_P and in payoff inequality. The maximum likelihood estimates and the implied escalation rate and disagreement rate predictions are shown in Table 7 and 8, respectively.[33]

Notice that the best-fitting logit-QRE correctly predicts nearly all of the (directional) comparative statics predictions (see Tables 7 and 8).

RESULT 7 *The best-fitting logit-QRE correctly predicts that*

- *the disagreement rate is significantly lower in HIGH than in BASE and LOW,*
- *the escalation rate...*

[32] Regarding conflict escalation, both the subgame perfect Nash equilibrium and the logit-QRE correctly predict that there is more escalation in the BASE condition than in the HIGH and LOW conditions. The logit-QRE is the more accurate of the two predicting a 71% (actually 86%) escalation rate in the BASE condition and a 49% in the other two conditions (actually 60% in the HIGH condition and 57% in the LOW condition) where it is suboptimal to escalate.

[33] In Table 8, we consider a model where preference parameters are zero at the negotiation table and positive at the escalation stage (see Cox et al. (2008), for instance). If one used a model with a model where preference parameters of the provoker are identical at the negotiation stage and at the escalation stage, the maximum likelihood parameters are $\mu = 3.33$, $\lambda = 0.2$ and $\gamma = 0.95$. The comparative statics predictions are unaltered in this case. Notice that QRE is not invariant to linear payoff transformations. We calculate the QRE prediction using the Euro-termed earnings as our numeraire.

– *is higher in RISKY than in CERTAIN,*

– *is higher in BASE than in HIGH and LOW,*

– *is highest in HIGH RISKY and lowest in LOW CERTAIN.*

These results are in line with Prediction 7. The subgame perfect Nash equilibrium with self-interested agents fails to predict these empirical comparative statics, apart from the prediction that BASE escalation rate is higher than the escalation rate in HIGH and LOW. As explained in Section 3, the subgame perfect Nash equilibrium disagreement rate is 0% in BASE and 100% in HIGH and LOW. The subgame perfect equilibrium predictions merely look at risk-neutral self-interested parties' optimal decisions without considering how strong a preference each party has for the preferred action; in order to have the parties reaching a settlement, conflict escalation has to be profitable for the provoker to create a credible threat. We do not observe the predicted extreme disagreement rates, and moreover even the empirical comparative statics are against the subgame perfect Nash equilibrium prediction: there is more disagreement in BASE condition than in HIGH condition and not vice versa as predicted by subgame perfection. The logit-QRE makes better predictions and even captures the comparative statics between BASE, HIGH, and LOW.

5 Discussion and concluding remarks

We study settlement negotiations and the decisions to escalate conflict if negotiations fail in an incentivized non-framed laboratory experiment. The provoker party deciding on escalation is an underdog earning less than the opposing defender side at all impasse outcomes, whether she escalates the conflict or not. At some of the escalation outcomes, the disadvantageous inequality is smaller in magnitude than at others. In line with subgame perfect Nash equilibrium under self-interest, escalation rates are higher when it is optimal to escalate than when not. Yet, contrary to the predictions of risk aversion, we find that escalation rates are higher when escalation outcomes are random rather than certain. We also find that increasing the cost of escalation is less effective in reducing the escalation rate than lowering the underdog's probability of winning when both changes are calibrated to reduce escalation incentives by the same magnitude for a self-interested risk-neutral party. Moreover, the agreement rate in the settlement negotiations is higher when escalation costs are higher compared to the case with lower winning probability. Thus there is a positive effect of higher escalation costs on efficiency at the negotiation stage, but a negative one at the escalation stage.

We find empirical support for the inequity-as-loss hypothesis where the reflection effect triggers risk-loving escalation patterns. Moreover, when embedded in a quantal response equilibrium, the proposed inequity-as-

loss model not only captures the escalation patterns but also organizes the decisions of the two parties at the negotiation stage.

In addition to our inequity-as-loss hypothesis, there are yet two alternative candidate explanations that suggest a higher escalation rate when escalation outcomes are risky, but these explanations are not consistent with our empirical evidence. First, people tend to hold illusions of controlling entirely aleatory events and being able to turn them in their favor (Langer, 1975)[34], and situations where favorable and non-favorable outcomes are salient are particularly likely to conceive such illusion of control (Thompson et al., 1998). In our case, the provokers' illusions about the random escalation outcome condition may have strengthened their faith in getting a favorable outcome.[35] Illusion of control suggests that there should be more escalation when escalation outcomes are aleatory: only when escalation outcomes are random can subjects hold an illusion of controlling the escalation outcomes. Illusion of control does not, however, predict that the escalation rate would be highest in the HIGH condition with risky escalation outcomes and lowest in LOW with certain outcomes. If it discriminates at all, it rather predicts the highest rate in the LOW condition with risky outcomes where the net gains are higher than in the HIGH condition.[36] Thus, to the extent that our data suggests the opposite, the evidence goes against this explanation.

A second alternative explanation holds that small-probability events are overweighted in human estimation of the likelihood of uncertain events. Guthrie (2000) extending Rachlinski's (1996) experimental analysis finds that small probability context reverses the choice patterns implied by standard loss aversion, so that choices appear risk loving (risk averse) in gains domain (loss domain) when the probability of winning (losing) is small (see also Harbaugh et al. 2002; 2010). Thus in our LOW condition the winning event might receive a higher weight in subjects' minds making the prospect of conflict escalation look overly favorable, while the same effect should be absent or weaker in the HIGH condition, since the chances of winning and losing are more equal. Therefore, probability overweighting predicts highest escalation rates in the risky LOW condition, but again our evidence goes against this explanation.[37]

The observed behavioral patterns cannot be explained by procedural fairness theories, either. The experimental data of Bolton et al. (2005) for instance, illustrates that although expected equality also matters for people when rejecting ultimatum offers, it is less influential than when equality can be generated with

[34] Conclusions of Langer's (1975) series of experiments have been confirmed later in a number of follow up studies. Notice also that since the escalation outcome probabilities and winning shares are publicly known, there is little room for self-serving biases about the outcomes that could bring about conflict or excessive escalation (Babcock and Lowenstein, 1997).

[35] To keep check of the illusion, our instructions explicitly emphasized that the outcome draw is a fully computerized random draw. Yet, the literature tells us that the phenomenon stands firm even when odds for winning are explicitly given (Thompson et al., 1998).

[36] See Alloy and Abramson (1979) or Dunn and Wilson (1990) for the vividness argument in illusion of control.

[37] The safe "no litigation"-option in our experiment does not yield the expected value of the escalation lottery (from the social comparison perspective). This is where the choice task considerably differs from Harbaugh et al. (2002; 2010) and thus probability overweighting should indeed lead to higher escalation of conflict according to probability overweighting. See the online appendix for the formal argument.

certainty.[38] Thus procedural fairness should predict less escalation under risky outcomes, whereas we find that the underdogs are more willing to escalate conflict when escalation outcomes are risky.

Spitefulness (Levine, 1998) and reciprocity (Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fishbacher, 2006; Cox et al. 2007) provide further potential explanations for why the provoker is interested in reducing the payoff difference between herself and the defender. These models also correctly predict the pattern that there is more escalation in our main treatments with negotiations than in the additional treatments without negotiations.[39] Crucially, reciprocity models fail also to account why escalation is more common when escalation outcomes are risky than when they are certain.

The predictions of self-interested risk aversion also run counter to our empirical patterns. Such models predict that the escalation rate is lower when escalation outcomes are stochastic (see Holt and Laury, 2002, for instance). We find the exact opposite: there is more escalation in the conditions where trial outcomes are stochastic. We are thus left with the inequity-as-loss explanation.

Our results may be accentuated by a potential framing effect driven by the fact that in our instructions, we use the word "sharing" in the negotiation context. This wording may trigger an idea that resources ought to be divided equally. If this is the case, the proposals and minimal acceptable offer thresholds may be shifted closer to equal shares, and the failure to agree in the equal terms indicated in the bargaining strategies may trigger conflict escalation more easily than if less normative language was used in the instructions (see Hoffman et al. 1994, 1996, for instance). This would also potentially accentuate the differences in escalation rates between the treatments with and without negotiations.

Another potential shortcoming is the fact that, despite the strategy method approach, our study may not fully account for the endogeneities and selection effects that may confound the escalation data. To see this, notice that any differences in disagreement rates between the six conditions also induce differences in the expected circumstances under which the escalation choices are payoff relevant. However, due to the strategy method, at the time of deciding whether to escalate or not, it is unknown whether there will be disagreement or not. Therefore, we have escalation choice data from every round of play and every pairwise match of participants. Our method thus perhaps limits selection effects but does not remove them altogether.[40]

The key characteristics of our experimental setup match with those of legal disputes, wage negotiations, or even international conflict. Yet one should acknowledge that the empirical data comes from laboratory

[38] They study subjects in simplified ultimatum games where the pie can only be shared in two asymmetric ways: 80% for proposer and 20% for responder or 20% for proposer and 80% for responder. They found that subjects were more willing to reject proposals favoring the proposer if the proposer had an alternative option to propose a lottery over the same unequal outcomes but with equal expected payoffs. The responder could decide whether to reject or accept that lottery without knowing its realization. Rejection led to zero payoffs for each side with certainty. Yet, the rejection rate of the proposal favorable to the proposer was even higher when there was a sure fifty-fifty split alternative available.

[39] The positive coefficient of Nego variable shown in Table 3 in the online appendix.

[40] Dal Bo et al. (2010) suggest a method for controlling such effects entirely but we did not apply their method in this paper.

experiments with students. Certainly, one should be cautious about extrapolating social preference evidence from laboratory to the field. Yet, the laboratory serves as a means of isolating and identifying factors that can and should be further scrutinized in the field (List, 2009). The reputational bargaining literature (Abreu and Gul, 2000; Compte and Jehiel, 2002; Fanning, 2016), where reputation for obstinacy provides an advantage to an institutional negotiator, suggests a reason why the results of a laboratory experiment on negotiations and conflict may be particularly likely to replicate in the field much the same way as there is evidence of reputational concerns amplifying the other-regarding behavioral tendencies in experimental labor markets (Fehr, Brown and Zehnder, 2009; Fehr, Goette and Zehnder, 2009). If the observed laboratory patterns constitute typical human behavioral tendencies in such a context, then they may be optimally mimicked in repeated real settings. Laboratory evidence in support of such reputation building in the lab is provided by Embrey et al. (2015) but field experimental evidence is still non-existent.

To which extent are professional negotiators capable of taming down instinctive reactions to sensation of frustration and injustice, or have been selected so as to lack that tendency, or do they mimc the behavioral reactions in order to build reputation for obstinacy? If not all negotiators are capable and willing to hold back the instinctive reactions, then there is a wide range of applications for which novel policy insights become relevant: how to modify institutions so as to curb excessive litigation caused by behavioral tendencies in patent cases for instance.[41] Similarly, firms' managements should be alert to behavioral biases potentially present in employees, as pointed out by Armstrong and Huck (2010) and Stucke (2014). An executive negotiating a merger may be influenced by the prospective personal incentives to close the deal at a price suboptimal to the firm. Such outcomes can be expected in particular in situations where the incentives set for the executive are unrealistically demanding placing her in a loss-frame against her target compensation or if the compensation of the executive relative to her peers is relatively low unless she receives the bonuses granted for closing the deal. Indeed, our results have particular relevance to firm management as they help correct badly designed incentive schemes that trigger too much risk taking by the employees compared to what would be best for the firm.

Ultimately, further evidence from artefactual, framed and natural field-experiments (Harrison and List, 2004) with exogenous variation in outcome probabilities and costs of escalation is needed to confirm the tendencies identified in our conventional lab experiment. There is an abundance of naturally-occurring happenstance data from legal disputes, for instance. Yet, in that data settled cases are under-represented because a large share of cases are settled out of court perhaps even before the case is filed and thus not

[41] Patent litigation is an especially relevant example, as the remedies for infringement include "reasonable royalties", and royalty payments for standard essential patents typically need to be "fair, reasonable, and non-discriminatory". Such language combined may allow for the formation of subjective beliefs of what "fair" and "reasonable" mean, which can lead to incongruous social value comparisons between the patent holder and the alleged infringer and result in excess litigation.

observed, whereas a laboratory experiment fully avoids this selection bias.[42]

References

Abreu, D., Gul, F. 2000. Bargaining and Reputation. *Econometrica* 85-117.

Alloy, L. B., Abramson, L. Y. 1979. Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser? *Journal of Experimental Psychology: General* 108, 441-485.

Anbarci, N. Feltovich, N. 2013. How sensitive are bargaining outcomes to changes in disagreement payoffs? *Experimental Economics* 16, 560-596.

Andersson, O., Holm, H. J., Tyran, J. R., Wengström, E. 2015. Deciding for Others Reduces Loss Aversion. *Management Science*, forthcoming.

Armstrong, M., Huck, S. 2010. Behavioral Economics as Applied to Firms: A Primer. Competition Policy International 6, 3-45.

Babcock L., G. Loewenstein 1997. Explaining Bargaining Impasse: The Role of Self- Serving Biases. *Journal of Economic Perspectives* 11, 109-126.

Bardsley, N. 2008. Dictator Game Giving: Altruism or Artefact? *Experimental Economics*, 11, 122-133.

Bault, N., Coricelli, G., Rustichini, A. 2008. Interdependent Utilities: How Social Ranking Affects Choice Behavior. *PLOS One* 3, 1-10.

Bellemare, C., Kröger, S., and Van Soest, A. 2008. Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica* 76, 815-839

Bolton, G. E. 1991. A Comparative Model of Bargaining: Theory and Evidence. *American Economic Review* 71, 1096-1136.

Bolton, G. E., and Ockenfels, A. 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* 90, 166-193.

Bolton, G. E., Brandts, J., Ockenfels, A. 2005. Fair Procedures: Evidence from Games Involving Lotteries. *Economic Journal* 115, 1054-1076.

Bolton, G. E., and Ockenfels, A. 2010. Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment. *American Economic Review* 100, 628-633.

Brekke, K., Ciccone, A., Heggedal, T.-R., Helland, L., 2016. Reference points in sequential bargaining: Theory and experiment. Manuscript, 2016.

[42]Consequently licenses from public sources should be interpreted with caution, for example the license agreements filed to U.S. Securities and Exchange Commission.

Brennan, G., González, L. G., Güth, W., Levati, M. V. 2008. Attitudes toward Private and Collective Risk in Individual and Strategic Choice Situations. *Journal of Economic Behavior and Organization* 67, 253-262.

Charness, G., and Rabin, M. 2002. Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* 117, 817-869.

Compte, O., Jehiel, P. 2002. On the Role of Outside Options in Bargaining with Obstinate Parties. *Econometrica* 70, 1477-1517.

Costa-Gomes, M. A., Zauner, K., 2001. Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis. *Games and Economic Behavior* 34, 238-269.

Cox, J., Friedman, D., Gjerstad, S. 2007. A Tractable Model of Reciprocity. *Games and Economic Behavior* 59, 17-45.

Dal Bo, P., Foster, A., Putterman, L. 2010. Institutions and Behavior: Experimental Evidence on the Effects of Democracy. *Amerian Economic Review* 100, 2205-2229.

Dechenaux, E., Kovenock, D., and Sheremeta, R. M. 2015. A Survey of Experimental Research on Contests, All-pay Auctions and Tournaments. *Experimental Economics* 18, 609-669.

Dufwenberg, M., and Kirchsteiger, G. 2004. A Theory of Sequential Reciprocity. *Games and Economic Behavior* 47, 268-298.

Dunn, D. S., Wilson, T. D. 1990. When the Stakes Are High: A Limit to the Illusion-of-control Effect. *Social Cognition* 8, 305-323.

Eisenkopf, G., and Teyssier, S. 2013. Envy and Loss Aversion in Tournaments. *Journal of Economic Psychology* 34, 240-255.

Embrey, M., Frèchette, G. R., and Lehrer, S. F. 2015. Bargaining and Reputation: An experiment on Bargaining in the Presence of Behavioural Types. *Review of Economic Studies* 82, 608-631.

Falk, A., and Fischbacher, U. 2006. A Theory of Reciprocity. *Games and Economic Behavior* 54, 293-315.

Fanning, J. 2016. Reputational Bargaining and Deadlines. *Econometrica* 84, 1131-1179.

Fehr, E., Brown, M., and Zehnder, C. 2009. On Reputation: A Microfoundation of Contract Enforcement and Price Rigidity. *Economic Journal* 119, 333-353.

Fehr, E., Goette, L., and Zehnder, C. 2009. A Behavioral Account of the Labor Market: The Role of Fairness Concerns. *Annual Review of Economics* 1, 355-384.

Fehr E., Schmidt, K. M. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114, 817-868.

Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171-178.

Friedman, M. 1953. The Methodology of Positive Economics. . Reprinted in Hausman, D. (ed.) *Philosophy of Economics: An Anthology*. 2nd edition, Cambridge University Press, UK.

Gamba, A., Manzoni, E., and Stanca, L. 2016. Social Comparison and Risk Taking Behavior. *Theory and Decision*, *forthcoming*.

Goeree, J., Holt, C. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *American Economic Review* 91, 1402-1422.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. 2016. *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press, NJ, USA.

Greiner, B. 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1, 114-125.

Guthrie, C. 2000. Framing Frivolous Litigation: A Psychological Theory. *University of Chicago Law Review*, 163-216.

Haisley, E., Mostafa, R., Lowenstein, G. 2008. Myopic Risk-seeking: The Impact of Narrow Decision Bracketing on Lottery Play. *Journal of Risk and Uncertainty* 43, 141-167.

Harbaugh, W. T., Krause, K., and Vesterlund, L. 2002. Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses. *Experimental Economics*, 5, 53-84.

Harbaugh, W. T., Krause, K., and Vesterlund, L. 2010. The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks. *Economic Journal* 120, 595-611.

Harrison, G. W., and List, J. A. 2004. Field Experiments. *Journal of Economic Literature*, 42, 1009-1055.

Herbst, L. Konrad, K., Morath, F. 2017. Blance of power and the propensity of conflict. *Games and Economic Behavior*, 103, 168-184.

Holt, C. A., Laury, S. K. 2002. Risk Aversion and Incentive Effects. *American Economic Review* 92, 1644-1655.

Jackson, M. O., and Morelli, M. 2011. The Reasons for Wars: an Updated Survey. In: C. J. Coyne and R. L. Mathers (eds.) *Handbook on the Political Economy of War*, 34.

Kahneman, D, Tversky A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 263-291.

Karagözoglu, E. Keskin, K. 2018. Endogenous Reference Points in Bargaining. *Mathematical Methods of Operations Research forthcoming*.

Kennan, J., and Wilson, R. 1989. Strategic Bargaining Models and Interpretation of Strike Data. *Journal of Applied Econometrics* 4(S1).

Kimbrough, E. Sheremeta, R. 2014. Why can't we be friends? Entitlements and the costs of conflict. *Journal of Peace Research* 51, 487-500.

Kimbrough, E. Sheremeta, R, Shields, T.. 2014. When Parity Promotes Peace: Resolving Con ict Between Asymmetric Agents. *Journal of Economic Behavior and Organization* 99, 96-108.

Konrad, K. A. 2009. *Strategy and Dynamics in Contests*. Oxford University Press, Oxford, UK.

Korobkin, R. B. 2002. Aspirations and Settlement. *Cornell Law Review* 80, 1-61.

Lacina, B., and Gleditsch, N. P. 2005. Monitoring Trends in Global Combat: A new Dataset of Battle Deaths. *European Journal of Population* 21, 145-166.

Langer, E. 1975. The Illusion of Control. *Journal of Personality and Social Psychology*, 32, 311-328.

Laury, S. K., and Holt, C. A. 2008. Payoff Scale Effects and Risk Preference under Real and Hypothetical Conditions. In: C. R. Plott and V. L. Smith (eds.) *Handbook of Experimental Economics Results* 1, 1047-1053.

Levine, D.K. 1998. Modelling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* 1, 593-622.

Linde, J., Sonnemans., J. 2012. Social Comparison and Risky Choices. *Journal of Risk and Uncertainty* 44, 45-72.

List, J. A. 2007. On the Interpretation of Giving in Dictator Games. *Journal of Political Economy* 115, 482-493.

List, J. A. 2009. Social Preferences: Some Thoughts from the Field. *Annual Review of Economics* 1, 563-579.

List, J. A., Shaikh, A. M. , Xu, Y. 2016. Multiple Hypothesis Testing in Experimental Economics. *National Bureau of Economic Research Working Paper* No. w21875.

Loewenstein, G. F., Thompson, L., Bazerman, M. H. 1989. Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology* 3, 426-441.

López-Vargas, K. 2014. Risk attitudes and fairness: Theory and experiment. *Department of Economics Working Paper, University of Maryland*, 6, 12.

McKelvey, R. D., Palfrey, T. R. 1998. Quantal Response Equilibria for Extensive Form Games. *Experimental Economics* 1, 9-41

Ostrom, B.J., Kauder, N., B., LaFountain, R. C. 2003. *Examining the Work of State Courts, 2002: A National Perspective From the Court Statistics Project*, National Center for State Courts, Williamsburg, VA, USA.

Rachlinski, J. 1996. Gains, Losses, and the Psychology of Litigation. *Southern California Law Review* 70, 113-185.

Robson, A. J. 1992. Status, the Distribution of Wealth, Private and Social Attitudes to Risk. *Econometrica* 60, 837-857.

Rohde, I., Rohde, K. 2011. Risk Attitudes in a Social Context. *Journal of Risk and Uncertainty* 43, 205-225.

Schlag, K. H., Tremewan, J., Van der Weele, J. J. 2015. A penny for your thoughts: A survey of methods for eliciting beliefs. *Experimental Economics* 18, 457-490.

Spier, K. E. 2007. Litigation. In: A. M. Polinsky and S. Shavell (eds.) *Handbook of Law and Economics 1*, pp. 259-342.

Stucke, M. E. 2014. How Can Competition Agencies Use Behavioral Economics? *Antitrust Bulletin* 59, 695-742.

Thaler, R. 1980. Towards a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization* 1, 39-60.

Thompson, S. C., Armstrong, W., Thomas, C. 1998. Illusions of Control, Underestimations, and Accuracy: A Control Heuristic Explanation. *Psychological Bulletin* 123, 143-161.

Trautmann, S. T., and Vieider, F. M. (2012). Social Influences on Risk Attitudes: Applications in economics. In: S. Roeser, R. Hillerbrand, P. Sandin, M. Peterson (eds.) *Handbook of Risk Theory* (pp. 575-600). Springer Netherlands.

Tversky, A., Kahneman, D. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.