Bilateral Conflict: An Experimental Study of Strategic Effectiveness and Equilibrium*

Charles A. Holt and Thomas R. Palfrey

December 19, 2022

Short/Running Title: Binary Conflict Experiment

Keywords: international conflict, terrorism deterrence, attacker-defender games, Nash equilibrium, quantal response equilibrium

---

Charles A, Holt, A. Willis Robertson Professor of Political Economy, Department of Economics, University of Virginia, Charlottesville, VA 22904-4182, cah2k@virginia.edu
Thomas R. Palfrey, Flintridge Foundation Professor of Economics and Political Science, HSS Division, California Institute of Technology, Pasadena, CA 91125 trp@hss.caltech.edu.

*We would like to acknowledge the thoughtful suggestions provided by Clement Bohr, helpful discussions with Michael Gibilisco, and the research assistance provided by Nicholas Anderson, Georgia Beazley, Courtney Carpenter, Jessica Ewing, Madeleine Green, Yilin Huang, Maggie Isaacson, Anna Kaplan, Sophia Lochner, Madison Smither, Erica Sprott, Allison Sullivan, Elin Woolf, and Max Zilvitis. Institutional Review Board approval for this research was obtained from the University of Virginia.

Abstract

Bilateral conflict involves an attacker with several alternative attack methods and a defender who can take various actions to better respond to different types of attack. These situations have wide applicability to political, legal, and economic disputes, but are particularly challenging to study empirically because the payoffs are unknown. Moreover, each party has an incentive to behave unpredictably, so theoretical predictions are stochastic. This paper reports results of an experiment where the details of the environment are tightly controlled. The results sharply contradict the Nash equilibrium predictions about how the two parties' choice frequencies change in response to the relative effectiveness of alternative attack strategies. In contrast, nonparametric quantal response equilibrium predictions match the observed treatment effects. Estimation of the experimentally controlled payoff parameters across treatments accurately recovers the true values of those parameters with the logit quantal response equilibrium model but not with the Nash equilibrium model.

The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the American Journal of Political Science Dataverse within the Harvard Dataverse Network at: https://doi.org/10.7910/DVN/TAOLXZ

Word Count: 10,000

I. Introduction

Many settings that arise in the study of politics and international relations involve strategic conflict between two opposing sides with asymmetric roles, where the defender takes a position to fend off the attacker's attempts to exploit vulnerabilities. A country facing an imminent attack must consider how to deploy forces across alternative points of attack, as in the classic Colonel Blotto game (Roberson 2006), and the invader must choose how best to allocate attack assets. A government facing an active domestic or foreign terrorist group must decide how to allocate security forces across multiple potential targets such as stadiums, theaters, markets, airports, pipelines, and schools (Powell 2007a, Powell 2007b, Sandler and Lapan 1988).

Applications of attacker-defender games in political science are not limited to situations of violent conflict, but encompass a broader range of environments, including electoral competition. A vulnerable centrist incumbent politician facing a possible electoral challenge from the left or the right must choose which policy position to adopt to fend off alternative types of challenges (Ansolobehere et al. 2001). Candidates who must win both a primary election and a general election face a similar strategic positioning problem (Brady et al. 2007, Gerber and Morton 1998).

Even though both sides may have a rich array of available actions in these examples, it is useful to consider broad classifications to isolate salient aspects of the problem for the design of a laboratory experiment. Military assaults may be either frontal or indirect, as with amphibious landings farther from the front. Candidate positions can be left or right. The games to be studied model the binary conflict as one where the attacker has two alternative strategies, while the defender has two defensive strategies, one of which is more effective against a particular attack.

The payoffs for this class of games are shown in Table 1, with the payoff for the defender (row player) listed on the left side of each cell. The attacker (column player) chooses between actions a_1 and a_2, which could represent a terrorist attack at site 1 or site 2. The defender allocates defensive resources between these two alternative attacks. The defense is most effective by matching the attack, and the attack is most effective against a "mismatched" defense. The attacker's payoff listed on the right side of each cell can be interpreted as the probability of a successful attack. This payoff is highest when the subscripts for the attack and defense actions do not match. For example, attack a_1 is more effective against defense d_2, where the positive difference $B-A$ represents the relative advantage of this attack against a mismatched defense.

An Attacker-Defender Game
with Choice Probabilities α for Attack a_1 and δ for Defense d_2
(defender payoff, attacker payoff)

	a_1 (α)	a_2 ($1-\alpha$)
d_1 ($1-\delta$)	$-A$, A	$-C$, C
d_2 (δ)	$-B$, B	$-D$, D

Besides the substantive importance of bilateral conflict situations, our study is also motivated by more general theoretical and behavioral questions. One key issue is the predictive value of Nash equilibrium – the workhorse theoretical paradigm for analysis of strategic interactions (Holt and Roth, 2004). The Nash equilibrium of these games, which is in mixed strategies, makes sharp predictions about how behavior responds to changes in the relative payoffs of different combinations of attacker and defender choices.

Nash predictions in binary conflict games, however, can be highly counter-intuitive, which raises questions about its usefulness as a model. For example, one natural question that one might hope a model could answer involves predictions about how the probability of using an attack depends on its effectiveness. Consider an increase in the effectiveness of the first attacker strategy, a_1, so that the probability of success strictly increases against both defenses. It is reasonable to expect that the probability of action a_1 would increase, but Nash equilibrium predicts that such an increase in effectiveness can have no effect at all, or may even go in the opposite direction and lead to a reduction in the use of the enhanced attack.

In spite of these implausible Nash predictions, there is a widely used modification of Nash equilibrium that incorporates a specific type of bounded rationality and generates intuitive predictions: *quantal response equilibrium*. (QRE, McKelvey and Palfrey, 1995;

Goeree, et al., 2016, 2020). This approach, which has been applied to the analysis of both experimental and field data, is a stochastic generalization of Nash equilibrium. It maintains the equilibrium assumption of rational expectations (i.e., players' beliefs about the distribution of their opponents' actions are correct). QRE formally incorporates stochastic choice via privately observed payoff disturbances, which can be modeled as payoff-responsive errors: Actions with higher expected utility relative to the unperturbed payoffs are chosen more frequently than actions with lower expected utility, as in logit and probit models in discrete choice analysis. In less formal language, players are unable to optimize perfectly, but small mistakes are more likely than large ones, which captures a range of behavioral phenomena such as misperception, miscalculation, or inattention.

Our experiment involves manipulating the payoffs by increasing the effectiveness of attack a_1 against either defense in specific ways. Nash equilibrium predicts *no increased use* of the enhanced attack a_1, whereas QRE can explain the increased attack rate that is observed in the experiment. Since both approaches are being used for structural empirical work in political science, our findings may help inform researchers who face a choice of whether to analyze data through the lens of Nash equilibrium or QRE. The findings are also suggestive to formal theorists about the potential benefits of analyzing their models using a QRE approach.

In addition to providing insights for this specific class of games and the potential for using QRE in formal modeling and in the structural estimation with field data, the experiment also illustrates three advantages of laboratory experiments that make them especially valuable for theory-testing exercises like this one.

One obvious advantage is that the payoffs of the game matrix are *controlled* in the laboratory, and hence known to the experimenter and all players, and are assigned exogenously in a manner that prevents unobserved nuisance variables from interfering with causal inference. In contrast, a comparable analysis of naturally occurring games would require estimation of all the relevant payoff parameters. Moreover, unobserved payoff heterogeneity creates further challenges for estimation with field data, and additional assumptions about the players' beliefs about payoff differences are required.

The second advantage concerns the *precision* of control. Because the experimenter can specify exact payoffs and subject payoff information, manipulation of payoffs to test comparative static predictions can be done with precision. In relevant naturally occurring data, for example athletic contests (Walker and Wooders 2001 and Chiappori et al. 2002), the effect of enhanced attack effectiveness on the *magnitude* of the payoff changes (for example, *B-A*) is difficult to control or estimate precisely.

The third advantage of a laboratory experiment over field data is *design*. In the laboratory, one can *choose* which of the payoff manipulations provide a powerful test of the hypotheses, e.g. when different theories make sharply opposing predictions. In contrast, with naturally occurring data, it is usually the case that the payoff manipulations are neither directly observable nor a choice variable for the researcher.

These advantages of *control, precision, and design* should not be interpreted as a call to replace the study of naturally occurring data with laboratory data. To the contrary, various data sources each have their specific advantageous features. Naturally occurring data, as well as data generated by field experiments, are typically analyzed in the specific context of a significant substantive or policy problem, enriched by important institutional details.

The next section presents the model and analyzes the Nash and QRE predictions for changes in the effectiveness of one attack option against either defense. Section III details the design of the laboratory experiment. The main features of the data and statistical tests for treatment effects are presented in Section IV. In section V, we report formal statistical estimation and model comparison. We first fit a parametric logit QRE model to the data to show that resulting estimates of choice frequencies produce a close match to the data, while the Nash equilibrium choice frequencies are far from the data in all but one treatment. Our second approach to model comparison is to estimate the payoff parameters in a game experiment *as if* they were unknown. This enables us to test whether the estimates based on the QRE and Nash models successfully recover those known values. The QRE model accurately recovers the concealed payoff parameters, while Nash equilibrium does not. Section VI circles back to the original motivation for the experiment in terms of applications to field data and policy relevance for the design of optimal defense strategies against terrorist threats.

II. The Model, Nash Equilibrium, and QRE

A. The Model

Consider the zero-sum game in Table 1 in the previous section, with the defender's payoff shown as negative on the left in each cell of the table. The assumption that the attacker action does better against the "wrong" defense (i.e., with mismatched subscript) is indicated by the two inequalities in (i) below. The game is only interesting when there are no dominated strategies, which is ensured by assuming that the highest attacker payoff for each decision is greater than the lowest payoff for the other one, as specified in (ii):

i) $B > A > 0$ and $C > D > 0$

ii) $C > A$ and $B > D$

B. Nash Equilibrium

The Nash equilibrium is unique and is in mixed strategies, i.e. a pair of choice probabilities between 0 and 1, one for each player. Let α denote the probability that Column chooses attack a_1 and let δ denote the probability that Row chooses defense d_2; these probabilities are shown in parentheses next to the corresponding action in Table 1. The attacker's expected payoffs for each action, as a function of the defender's mixed strategy, are: $E_{a_1}(\delta) = A(1-\delta) + B\delta$ and $E_{a_2}(\delta) = C(1-\delta) + D\delta,$ so the expected payoff difference is:

$$E_{a_1}(\delta) - E_{a_2}(\delta) \; = \; A - C \; + \; (C - D + B - A)\delta.$$

This expected payoff difference must be 0 in a mixed strategy Nash equilibrium, which determines the equilibrium probability δ^* for the defender:

(1) $\delta^* = \frac{C-A}{C-D+B-A}$ (Nash equilibrium probability of d_2).

Note that this expression has been conveniently organized so that the various parameter differences ($B-A$), ($C-A$), and ($C–D$) are positive by the initial inequality assumptions.

The equilibrium probability for the attacker is obtained similarly. The expected payoff difference for d_2 versus d_1, as a function of the attacker's mixed strategy, α, is:

$$E_{d_2}(\alpha) - E_{d_1}(\alpha) = (C - D) - (C - D + B - A)\alpha,$$

which yields:

(2) $\alpha^* = \frac{C-D}{C-D+B-A}$ (Nash equilibrium probability of a_1).

Now consider an exogenous change that increases the effectiveness of attack a_1, by raising *both* A and B by the same amount. Notice that these parameters only appear as a difference ($B–A$) in the denominator of (2). This yields the unintuitive prediction that equal *absolute* increases in the effectiveness parameters A and B for attack a_1 will have no effect on the Nash equilibrium probability of choosing that attack. This prediction is due to a countervailing increase in the equilibrium probability of defending against that attack.[1]

On the other hand, it is clear from the formula for α^* in (2) that *unequal* increases in both A and B can change the Nash equilibrium frequency of attack a_1. Our experiment also investigates the effects of *proportional* increase in A and B, to πA and πB, where $\pi > 1$. Such proportional increases in the effectiveness of attack a_1 cause an increase in the denominator of (2), so the Nash probability of attack a_1 actually *decreases*. As before, this seemingly unintuitive prediction is due to a countervailing increase in the equilibrium probability of defending against that attack.[2]

---

[1] In practice, it is unlikely that the increases or decreases in attack effectiveness are exactly the same for both attack methods in naturally occurring settings. However, the particular case of equal enhancements provides an ideal baseline setting for a controlled experiment in which the theoretical Nash prediction of a null effect is sharply different from an alternative theoretical approach based on a generalization of the Nash equilibrium, described in the next section.

[2] Powell (2007a) also considered a 2-site attacker-defender model, with the difference that the defender has a continuous choice of how to allocate a divisible defense asset. He uses the indifference requirement of a Nash equilibrium in mixed strategies to generate comparative static results about the effect of decreased defense effectiveness that "…may at first seem to be counterintuitive" (p. 531) but which are consistent

C. Quantal Response Equilibrium

With two decisions for each player, the QRE is specified by a pair of choice probabilities, α and δ, as functions of (equilibrium) expected payoff differences, in a manner that depends on the "precision" of choice behavior, denoted by $\lambda>0$. As precision is increased, behavior gradually approaches fully rational best replies as choice probabilities respond more sharply to expected payoff differences.

Formally, the Nash indifference equations that generated (1) and (2) are replaced by QRE choice probability equations. Let $Q(\Delta)$ be any continuous, non-negative, and strictly increasing *quantal response function* of the precision-weighted expected payoff difference, Δ, between actions a_1 and a_2 (or between d_2 and d_1):

(3) $\alpha \;=\; Q\big(\lambda[E_{a_1}(\delta) - E_{a_2}(\delta)]\big) \;=\; Q(\lambda[\,\boldsymbol{A} - C + (C - D + B - A)\delta\,])$

(4) $\delta \;=\; Q\big(\lambda[E_{d_2}(\alpha) - E_{d_1}(\alpha)]\big) \;=\; Q(\lambda[\,C - D - (C - D + B - A)\alpha\,])$

Q is assumed to satisfy:

iii) $\lim_{\Delta\to\infty} Q(\Delta) = 1,$ and $Q(-\Delta) = 1 - Q(\Delta)$ for all Δ.

The limit condition ensures that choices are approximately deterministic for extreme payoff differences or high precision. The second part of (iii) requires that the choice probabilities are not biased for one attack action over the other, and (from continuity) implies that actions are chosen with equal probability when Δ is 0. In applied work, Q is often assumed to be a logit function, which will be used in the section on estimation.

A QRE is any pair of choice probabilities, α^* and δ^*, that solve equations (3) and (4). The α and δ probabilities on the right sides of these equations can be interpreted as representing "beliefs" about the other's strategy. Rational expectations requires that these beliefs match the true equilibrium choice probabilities on the left sides of (3) and (4).

Notice that the left equation in (3) can be rewritten in terms of the inverse of Q:

---

with equilibrium adjustments to changed conditions. The reductions in defense effectiveness in the Powell model are roughly analogous to the increases in attack effectiveness in the our game.

(5) $\frac{Q^{-1}(\alpha)}{\lambda} = \boldsymbol{A} - C + (C - D + B - A)\delta\ .$

As precision goes to infinity, the left side of (5) approaches 0, so the equation reduces to the standard indifference equation (1) for the Nash equilibrium defense strategy, δ^*. Similarly, (4) reduces to equation (2) for large λ. Thus, the mixed Nash equilibrium is the limiting case of QRE for any Q. Typically, different functional forms Q will provide different QRE outcomes for finite values of λ, but the key results about the effect of changing the effectiveness of a particular attack method that are nonparametric in the sense that these qualitative predictions of QRE hold for all Q.

QRE Predictions for Equal Absolute Changes in Attack Effectiveness Parameters

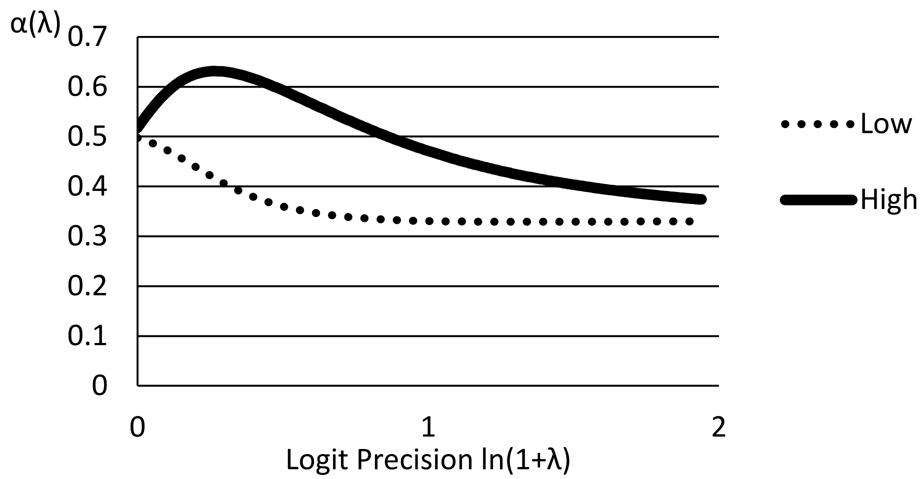
Even though equal absolute changes in the effectiveness parameters A and B will have no effect on the Nash equilibrium a_1 attack probability in (2), a quick inspection of the QRE equilibrium equations (3) and (4) indicates that equal changes in the A and B parameters will have an unambiguous effect on QRE predictions. There is an additional stand-alone A term on the right side of (3), which is highlighted in bold and directly affects the expected payoff difference for the attacker. With equal changes in A and B, this stand-alone term will generate a higher a_1 attack response for any given value of the defense choice probability, δ, on the right-hand side of (3). As a result, the QRE equilibrium a_1 attack rate, α, will *increase* when a constant is added to both A and B. This leads to Proposition 1 proof on p.2 of online appendix).

Proposition 1. *Under assumptions* (i), (ii), *and* (iii), *if the effectiveness parameters A and B increase by equal absolute amounts* to $A + \gamma$ and $B + \gamma$, *then the Nash equilibrium probability of attack* a_1 *is unchanged, but the QRE probability* α *of this attack method is increased. The QRE probability of defending against this attack* ($1-\delta$) *increases. This result holds for any quantal response function, Q and for any positive precision,* λ.

Proposition 1 holds under general conditions, and Figure 1 illustrates it for two of the payoff parameter sets used in the experiment. The figure shows the graph of the logit QRE a_1 attack rates as a function of λ (using a log scale) for two payoff profiles, $A = 2$,

B = 8, C = 7, D = 4, (Low effectiveness) and A = 6, B = 12, C = 7, D = 4, (High effectiveness with γ=4). For all positive values of λ, QRE predicts the "intuitive" pattern, with higher a_1 effectiveness resulting in higher a_1 choice frequencies.

Figure 1. Logit QRE a_1 Attack Probability

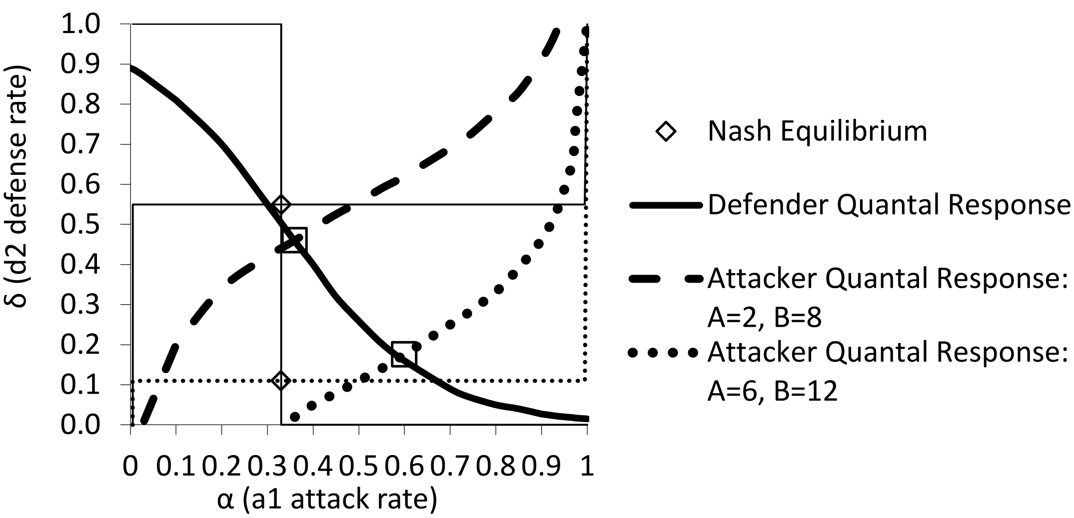


Note: The lines show the locus of QRE a_1 attack frequencies as precision increases from left to right, with attack effectiveness reduced in equal absolute amounts from High (solid line) to Low (dotted line.)

Notice that both curves in Figure 1 converge to the same Nash equilibrium mixed strategy of α = 0.33 as precision increases on the right side of the figure, as implied by the theory. However, the QRE correspondence exhibits the intuitive effect for a_1 attack proportions to be higher, *for all values of* λ for the High payoffs (solid line) compared with the Low payoffs. Furthermore, the convergence is from above for both treatments: i.e., QRE predicts $\alpha > 0.33$ for both treatments (or approximately equal to 0.33 for very large values of λ).

Figure 2 illustrates, in a different way, the equilibrium effects of equal absolute shifts in the attack effectiveness parameters A and B, by graphing the best response correspondences of the attacker and defender, and the analogous quantal response functions. The figure is constructed with α on the horizontal axis and δ on the vertical axis. For the Low payoffs (A = 2, B = 8, C = 7, D = 4), the attacker is indifferent between the two attack methods at $\delta = 0.55$, and so every mixture is a best response at that point, represented by the horizontal solid line at a height of $\delta = 0.55$. Similarly, the defender is indifferent between the two defense methods (and any mixture) when the a_1 attack is used with probability 0.33, which results in a vertical defender best response line at that point. The resulting best response graphs are step functions that intersect at $\alpha = 0.33$ and $\delta = 0.55$, the Nash equilibrium, marked by the solid diamond in the figure. Quantal response functions are the smooth curves shown in the figure for a logit specification ($\lambda = 0.7$). For the Low payoff profile, the attacker and defender quantal responses are shown as the curved upward sloping and downward sloping solid curves, respectively, which intersect at the Logit QRE slightly lower and to the right of the Nash equilibrium, marked by a solid square.

Figure 2. Effects of Equal Absolute Increases in Attack Effectiveness



Note: The mixed Nash equilibrium points (diamonds) are at intersections of sharp best response lines (straight or dotted). Quantal response equilibria (squares) are at intersections of the curved quantal response lines. The increase in attack effectiveness shifts the attacker quantal response line to the right, from dashed to dotted, which increases the predicted attack rate (square at lower right). The lower Nash equilibrium diamond for the dotted sharp best response line, in contrast, does not shift to the right.

Next consider the effect of equal absolute increases in A and B, from 2 and 8 to 6 and 12, respectively. The defender's quantal response function and best response curves depend only on the difference, $B-A$. Therefore both are invariant to equal *absolute* changes in A and B. In contrast, we can see from equation (3) that the attacker's quantal response function (and best response correspondence) shifts down in response to equal absolute increases in A and B, since α depends on A directly, as well as $A-B$. The shifted quantal response function is shown in the figure by the dotted upward sloping curve. This shift moves the QRE intersection down and to the right, as indicated by the square, so the QRE a_1 attack probability α increases as a result. Notice that this effect does not depend on the functional form of the quantal response function Q, as Proposition 1 established that the comparative static predictions are nonparametric.

For the Nash equilibrium analysis, the increased effectiveness of a_1 causes the attacker best response line to shift down (dotted horizontal line), but this does not change the Nash equilibrium probability of a_1 (marked by the diamond). This invariance is due to the fact that the vertical solid line at $\alpha = 0.33$ representing the indifference point for the defender is unaffected by equal increases in A and B, so the Nash equilibrium value of α is unchanged.

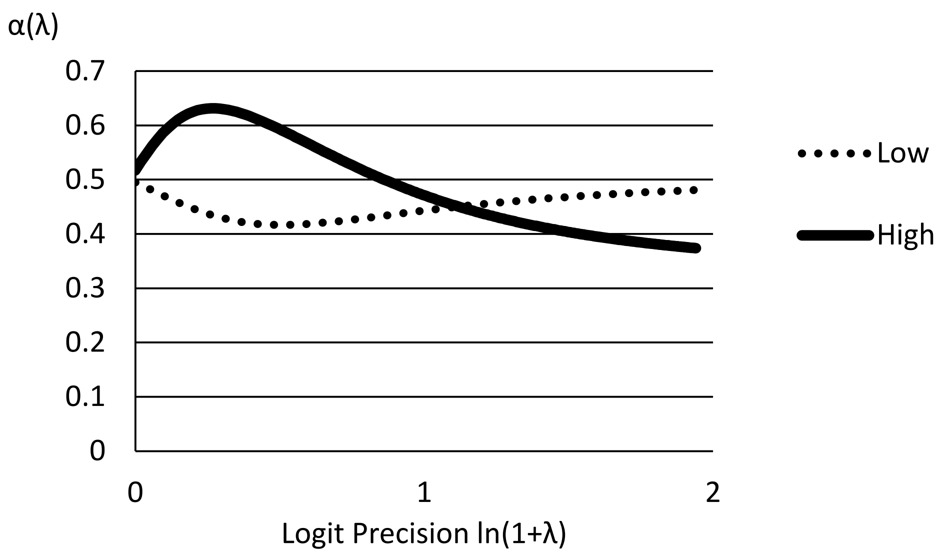
QRE Predictions for Equal Proportional Changes in Attack Effectiveness

As shown earlier, a *proportional increase* in attack effectiveness actually *decreases* the Nash equilibrium probability of this attack. For very high levels of precision, the QRE equilibrium necessarily converges to Nash equilibrium, so some QRE predictions would share this unintuitive prediction. However, this is not the case for a wide range of plausible lower precisions, as can be seen in Figure 3, which compares the logit QRE attack rates as a function of λ for two payoff profiles that are used in the experiment: $A = 3$, $B = 6$, $C = 7$, $D = 4$, (low effectiveness) and $A = 6$, $B = 12$, $C = 7$, $D = 4$, (high effectiveness). Except for very high values of λ, QRE predicts the "intuitive" pattern, with higher effectiveness resulting in higher a_1 attack proportions.

This effect of equal proportional changes in A and B shown for the example in Figure 3 is a general property of *any* QRE (not only logit), which leads to Proposition 2 (proof on pp.2-3 of online appendix):

Proposition 2. *Under assumptions* (i), (ii), *and* (iii), *if the* a_1 *attack effectiveness payoffs, A and B, increase by equal proportional amounts* (*to* πA *and* πB, *where* $\pi > 1$), *then for any quantal response function Q satisfying (iii), there exists a positive precision parameter* λ_0 *such that for* $0 < \lambda < \lambda_0$, *the* QRE a_1 *attack rate with high effectiveness is higher, while the analogous Nash attack rate is lower. Furthermore, the frequency of defending against* a_1 *increases in* π *for all values of* λ.

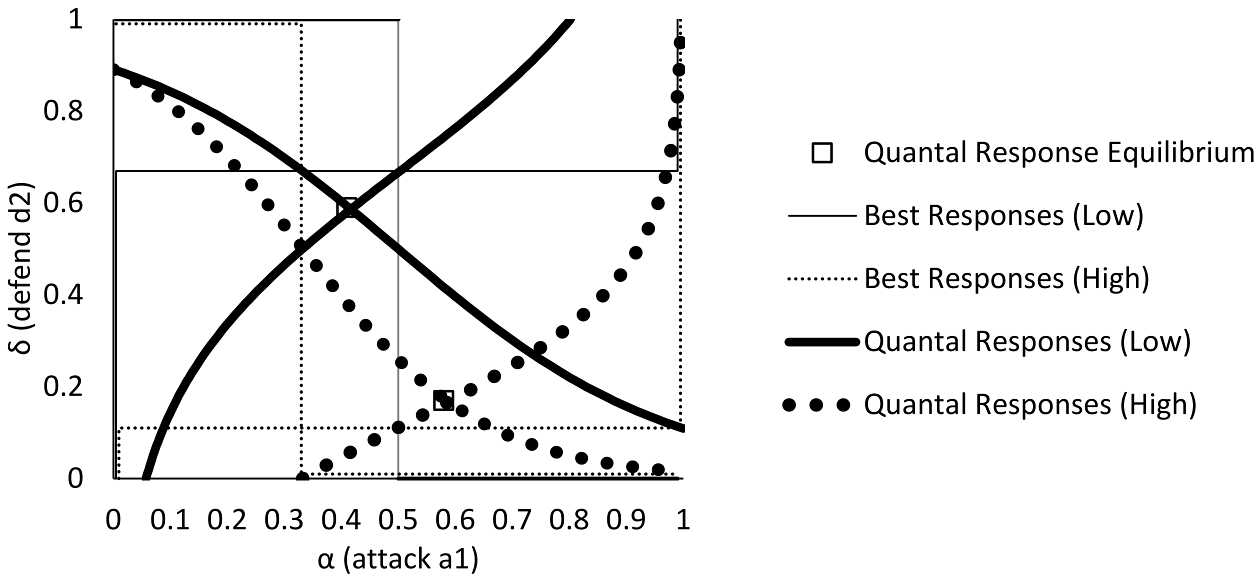
Figure 3. Logit QRE a_1 Proportions



Note: The lines show the QRE locus of predicted a_1 attack proportions as precision increases from left to right, with attack effectiveness reduced in equal proportions from High (solid line) to Low (dotted line).

Figure 4 shows the intuition behind the much different Nash and logit QRE predictions for changes in attacker and defender behavior resulting from an equal proportional increase in both *A* and *B*. The change induces a leftward shift of the downward sloping defense quantal response function (to the dark dotted curve) and a rightward shift of the upward-sloping attacker quantal response function (to the light dotted curve). The combined effect of both shifts is to reduce the QRE d_2 defense rate, δ, and *increase* in the QRE a_1 attack rate. In contrast, the Nash predictions for this treatment change are indicated by the switch from the solid, straight-line best responses in Figure 4 to the dotted-line best responses, resulting in a *decrease* in the Nash equilibrium a_1 attack rate from 0.50 to 0.33.

Figure 4. Effects of Equal Proportional Increases in A and B



Note: The sharp lines are best responses, the curved solid lines are quantal responses for low attack effectiveness, and the curved dotted lines are quantal responses for high attack effectiveness. The quantal response equilibrium square shifts to the right (with higher predicted attack rates). In contrast, the Nash equilibrium at the intersection of the sharp dotted lines shifts to the left (with lower predicted attack rates).

III. Experimental Design and Procedures

The experimental design, summarized in Table 1 uses the three payoff functions analyzed in the previous section. The design on the left side of the table involves *equal absolute changes* in the two attack effectiveness parameters, *A* (against defense d_1) and *B* (against defense d_2). The high effectiveness treatment is shown in the top row of Table 1, with a parameter set for *A*, *B*, *C*, *D* of 6, 12, 7, 4. The low absolute effectiveness treatment (henceforth called LowAb) is shown in the bottom row on the left side, with a reduction of both *A* and *B* by 4 in each case, i.e. from 6 and 12 to 2 and 8, respectively, resulting in no change of the Nash equilibrium a_1 attack rate of 0.33. The low proportional effectiveness treatment (henceforth called LowProp) is shown in the bottom row on the right side, with both *A* and *B* reduced by 50% from 6 and 12 in the top row to 3 and 6, resulting in an *increase* in the Nash equilibrium a_1 attack rate from 0.33 to 0.50.

Table 1. Treatments, Predictions, and Data Averages for Equal Absolute or Equal Proportional Changes in the Effectiveness of Attack a_1

	Equal Absolute Change		**Equal Proportional Change**	
	A, B, C, D	Nash a_1 Rate	A, B, C, D	Nash a_1 Rate
High Effectiveness of Attack a_1	6, 12, 7, 4	0.33	6, 12, 7, 4	0.33
Low Effectiveness of Attack a_1	2, 8, 7, 4	0.33	3, 6, 7, 4	0.50

Note: The table lists payoff parameters A, B, C, and D for the Attacker Defender Game, along with the Nash equilibrium a_1 attack rates for each of the four treatment cells. Low Effectiveness treatment in the bottom row of the table corresponds to relatively low attacker payoffs (A and B) for attack a_1, and the High treatment in the top row corresponds to high attacker payoffs. The switch from low to high attack effectiveness involves equal absolute changes on the left side and equal proportional changes on the right.

The experiment was conducted with payoffs (in dollars) reduced by a factor of 10, e.g. from 6 to 0.60, etc. We also added a fixed amount of 1.20 to all defender payoffs to avoid losses. These adjustments have no effect on the Nash equilibrium predictions in (1) and (2). We will continue to discuss the treatments in terms of integer amounts, e.g. 6, 12, 7, 4, which now refer to 10-cent payoff units.

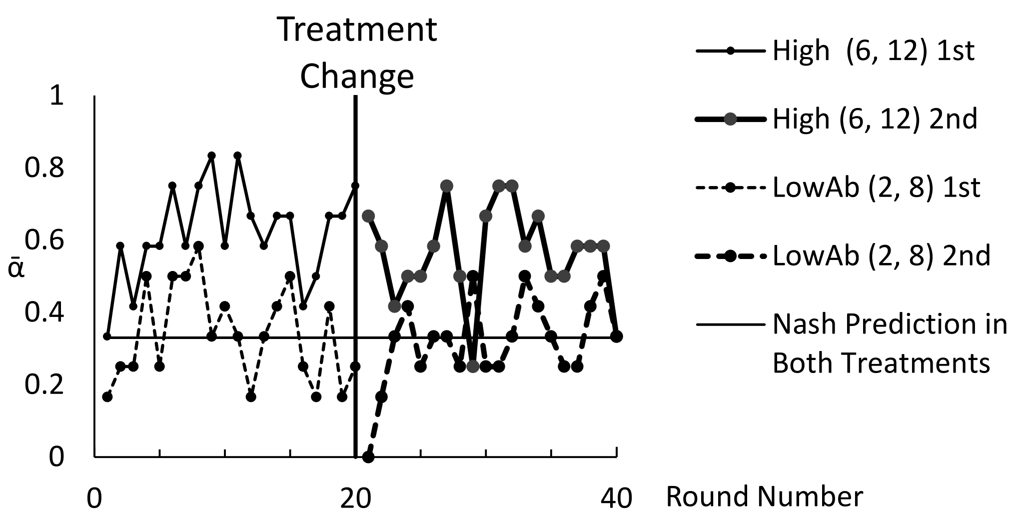
We conducted four sessions (48 subjects) with the equal absolute change design on the left side of Table 1, and another four sessions (48 subjects) with the equal proportional change design on the right side. Each session consisted of 20 rounds for each of the two attack effectiveness treatments (High and either LowAb or LowProp). Roles stayed the same (attacker or defender) and matchings were fixed, both within each 20-round part and across parts. Both treatment orders (Low-High and High-Low) were used, with Low-High in half of the sessions and High-Low in the other half. Subjects were recruited from the University of Virginia student population and were paid $6 plus all earnings. Total earnings averaged $30 for a session that lasted about 45 minutes, which included the reading of instructions aloud. The experiment was run with the publicly available online V*e*conlab software. The game was presented with minimal context, i.e.

with decisions labeled Top or Bottom and Left or Right. Instructions from one of the treatments are in online appendix p.6.

IV. Results

Figure 5 shows the time series of a_1 attack frequency averages ($\bar{\alpha}$) for the first paired treatment that changes the attack effectiveness parameters *A* and *B* by equal absolute amounts. This change does not affect the Nash equilibrium prediction for the a_1 attack rate, which is 0.33 for both treatments, as indicated by the horizontal dark line at that level. The solid lines that connect data average points in the figure are averages for the High treatment, and the dashed lines are for the LowAb treatment. The averages for

Figure 5. Average a_1 Attack Frequencies, $\bar{\alpha}$, by Round, with Equal Absolute Changes: Solid Lines for HighAb (6, 12), Dashed Lines for LowAb 2, 8)

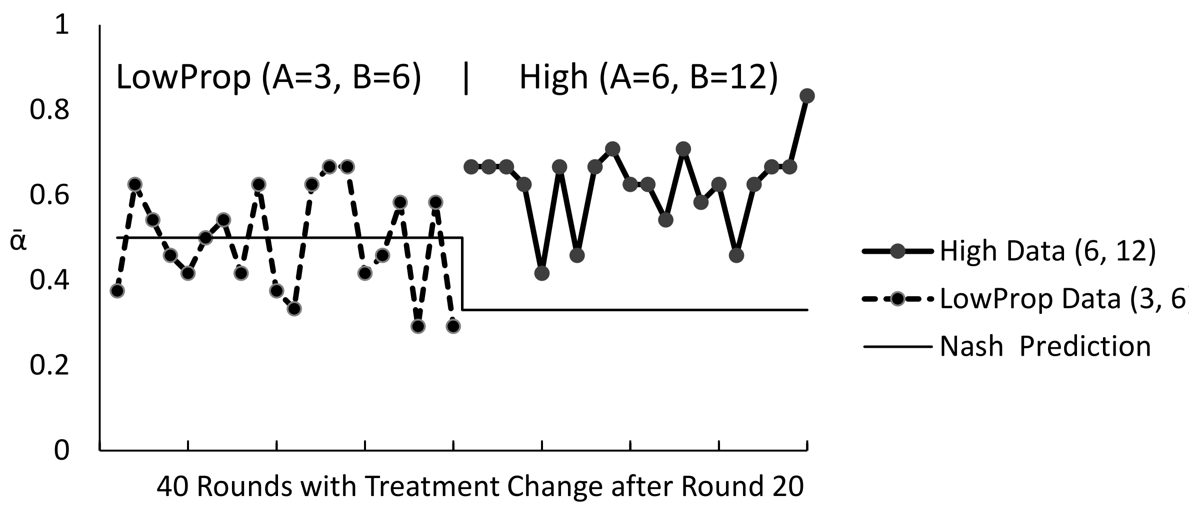


Note: The dashed lines connect data average dots for the LowAb treatment, and the solid lines connect data average dots for the High treatment. The Nash equilibrium prediction is represented by the flat line at 0.33 for both treatments. Data averages conform to Nash predictions in the LowAb treatment but are significantly higher High attack effectiveness treatment.

sessions with High in the first 20 rounds are displayed in gray, and averages for the reverse order are displayed in black. The black dashed line on the left shows that the observed a_1 frequencies center around the Nash prediction of 0.33 with low effectiveness for the first 20 rounds. The black solid line on the right indicates that a_1 attack frequencies almost doubled in the final 20 rounds after the treatment switch *for the same subjects and the same Nash prediction*. Average a_1 attack rates are greater in the High treatment, with an overall average of 0.59 (for both treatment orders), than in the Low treatment, with an overall rate of 0.33. Moreover, there are no clear sequence or trend effects.

Figure 6 shows a similar plot of a_1 frequencies for the case of equal proportional changes in the *A* and *B* effectiveness parameters. The treatment change from the LowProp treatment (3, 6) to the High treatment (6, 12) reduces the Nash equilibrium value of α from 0.50 (horizontal line on left side of Figure 6) to 0.33 (horizontal line on right), or vice versa for the opposite sequencing of the treatments. Data averages for the 20 rounds in the LowProp treatment are shown by a dotted line on the left side, regardless of whether it was done first or second, and the 20 rounds of data with the High treatment are displayed on the right. Observed a_1 attack frequencies increased from 0.49 to 0.63 as *A* and *B* increase proportionally, which is the opposite of the Nash equilibrium downward shift at the midpoint of the figure.

Figure 6. Average a_1 Attack Frequency by Round, with Equal Proportional Changes: Dashed Line for LowProp (3, 6), Solid Line for High (6, 12)



Note: The dashed line connects data average dots for the LowProp treatment, and the solid line connects data average dots for the High treatment. The Nash equilibrium prediction is represented by the flat lines that drop from 0.5 in the LowProp treatment to 0.33 for the High treatment, a predicted effect that is not observed in the attack rate data.

If the enhanced effectiveness of the a_1 attack is anticipated by the defenders, then the incidence of the matched d_1 defense should rise and the incidence for d_2 should fall. Table 2 shows a breakdown of the overall frequencies of both attack (a_1) and defense (d_2) decisions, with a row of treatment averages in bold and a row of Nash predictions in italics. First consider the left side of the table, for the equal absolute change. When the *A* and *B* effectiveness parameters for attack a_1 increase by equal amounts in this manner, the targeted defense d_1 against this attack is predicted to increase, and the propensity δ for the other defense d_2 is predicted to decline sharply from 0.56 to 0.11, as shown in the bottom row on the left. The corresponding increase in $1-\delta$ is what causes the Nash equilibrium value of α for attack a_1 to stay constant at 0.33, even in the presence of the doubled effectiveness of this attack method. The data show a less dramatic decline in the d_2 frequency, from 0.41 to 0.19, coupled with the a_1 frequency that almost doubles, from 0.33 to 0.59, which leads to our first result.

Result 1a: An equal <u>absolute</u> increase in effectiveness parameters for attack a_1 (from 2 and 8 to 6 and 12) causes a significant and large (nearly doubled) increase in the observed frequency of this attack, despite the Nash prediction of no change.

Table 2. Treatments, Nash Predictions, and Observed Data Proportions

	Equal Absolute Changes		**Equal Proportional Changes**	
Treatment Payoffs	LowAb	HighAb	LowProp	HighProp
(A, B, C, D):	(2 8 7 4)	(6 12 7 4)	(3 6 7 4)	(6 12 7 4)
Observed Attack Rate α:	0.33	0.59	0.49	0.63
Nash prediction:	0.33	0.33	0.50	0.33
Observed Defense Rate δ:	0.41	0.19	0.66	0.17
Nash prediction:	0.56	0.11	0.67	0.11

Note: The table shows attack and defense rates observed in the data for each of the four treatment conditions, with Nash equilibrium predictions listed in the row below in each case.

Support: The observed overall treatment difference is about 26 percentage points (0.59 – 0.33 = 0.26). The a_1 frequency increased in the High treatment for all but 6 of 24 attacker-defender pairs, with 3 relatively small reversals and 3 ties with equal attack rates in each treatment. Since each pair interacted with each other in all 20 rounds, we use a matched-pairs test to calculate the proportion of permutations of treatment labels that yield a treatment difference as great as or greater than the observed difference of 0.26. The resulting *p*-value is 0.0004, so the null hypothesis of no effect is soundly rejected.

The increase in a_1 attack rates in the HighAb treatment is consistent with the observation that the d_1 defense against this attack was observed less frequently than predicted in the Nash equilibrium. The increase in a_1 attack effectiveness resulted in a decline in the d_2 defense frequency from 0.41 to 0.19, which is a substantially smaller decline than the predicted decline from 0.56 to 0.11, as shown in the bottom row of Table 2, left side. This difference is highly significant:

Result 1b: An equal <u>*absolute*</u> *increase in effectiveness parameters for attack* a_1 *(from 2 and 8 to 6 and 12) results in a tendency to over-defend against the other attack* a_2 *relative to the Nash prediction of a sharper reduction in* d_2 *defense rates.*

Support: The observed overall reduction in the d_2 defense from 0.41 to 0.19 is about 22 percentage points, which is much less than the predicted 45 percentage point reduction from 0.56 to 0.11. The d_2 frequency decreased by less than 45 points for all but 4 of 24 attacker-defender pairs, as can be verified by considering the d_2 defense rate changes shown in the rows of p.5 of online appendix. This difference is significant at $p < 0.002$, confirming the QRE prediction that the overall reduction in the d_2 defense is less than the change in Nash frequencies of d_2.

Next consider the results for the equal proportional change with doubled *A* and *B* parameters, shown on the right side of Table 2. The Nash equilibrium value of α no longer stays the same, but rather, declines from 0.50 to 0.33. In contrast to this predicted decline, the attack frequency α *increased* from 0.49 to 0.63 percent, a 14 point change, yielding an a_1 attack frequency that is nearly double the Nash equilibrium frequency of 0.33. The observed 0.14 increase is significant for a 2-tailed matched pairs permutation test ($p = 0.034$). This is summarized as our second result:

Result 2a: An equal <u>*proportional*</u> *increase in effectiveness parameters for attack a_1 (from 3 and 6 to 6 and 12) causes a significant increase in the observed frequency for this attack, despite the Nash prediction that the frequency will decline.*

The final result pertains to the defender responses:

Result 2b: An equal <u>*proportional*</u> *increase in effectiveness parameters for attack a_1 (from 3 and 6 to 6 and 12) does not result in a clear tendency to over-defend against the other attack a_2 relative to the Nash prediction.*

Support: The overall observed 0.49 reduction in the average d_2 frequency from 0.66 to 0.17 is less than the predicted 0.56 reduction (from 0.67 to 0.11) in a Nash equilibrium. Furthermore, the d_2 frequency decreased by less than the predicted 56 points for two-thirds of the fixed attacker-defender pairs, but this is not statistically significant from one-half using a 2-tail binomial test ($p = 0.152$).

Finally, note that the choice frequencies in the third column of Table 3 for both attackers and defenders in the LowProp treatment are almost exactly equal to the Nash predictions (and a null hypothesis of no difference cannot be rejected by standard statistical tests). Some intuition for this result is suggested by Figure 4, where the solid quantal response curves for the LowProp treatment intersect at a point that is close the Nash intersection of the straight best response lines for this treatment. This suggests that Nash predictions are not badly flawed when Nash and QRE predictions are close, but less accurate otherwise, which leads into the next section where statistically compare the performance of QRE and Nash equilibrium models in these games.

V. Two Model Comparisons of QRE and Nash

The first approach is to obtain a maximum likelihood estimate of the responsiveness parameter of the logit QRE model by fitting the attack and defense frequencies in the data, and then compare the likelihood of the fitted model with the likelihood of the Nash equilibrium frequencies. There is a 1-1 mapping between λ and the profile of six QRE choice frequencies (for two player roles and three payoff treatments). Therefore, it is straightforward to obtain a maximum likelihood estimate of λ given by solving for the QRE choice rates at each stage of a search over a range of precision values (Goeree et al. 2016, Ch. 6).

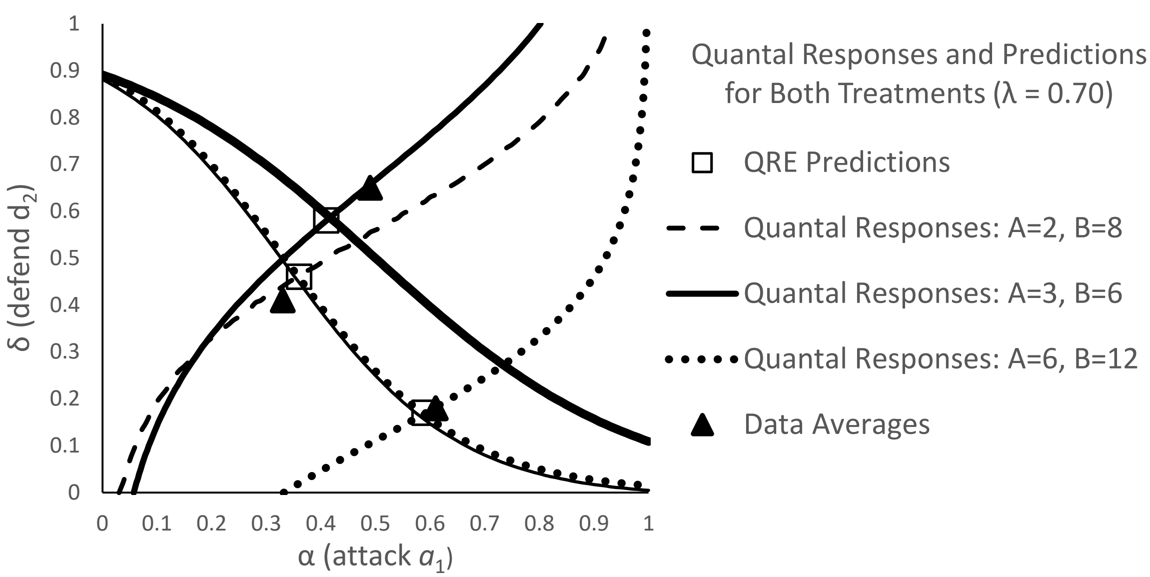
Our second method of model comparison is less standard in the analysis of experimental data and more commonly used in structural estimation of game theoretic models in empirical studies based on historical data or field observations. Such data sets often lack direct measurements of key structural parameters of the game-theoretic model applied to the empirical study, which in our case are the exact payoffs in the binary conflict game matrix. Hence the empirical analysis requires estimation of these parameters, under maintained assumptions about the equilibrium behavior of the agents. One can conduct the same estimation procedure with experimental data by concealing experimentally controlled parameters and treating them as if they were unknown. Merlo and Palfrey (2018) call this method of model validation "concealed parameter recovery" and used it to compare different behavioral models of voter turnout using experimental data. Here we use this method by concealing the two game payoff parameters (*C* and *D*) that are constant across our treatments and compare how well the QRE and Nash models recover those parameters.

A. Logit QRE Estimation

The previous analysis was based on qualitative data comparisons with patterns predicted by a Nash equilibrium or with general QRE comparative static predictions that hold non-parametrically. In this section, we report estimation results based on the parametric logit specification. The estimated value of the logit precision λ was constrained to be the same for all treatments. The quantal response functions (3) and (4) were rewritten as ratios of exponential expressions using the logit form: $Q(\lambda\Delta) = \frac{1}{1+\exp(-\lambda\Delta)}$, which resulted in two equations in the equilibrium attack and defense rates, α and δ, for a given logit precision λ. The equilibrium values of α and δ for each treatment were calculated numerically for all possible values of λ. The likelihood function is constructed as a product of each predicted choice probability (for each treatment) raised to a power that equals the number of times that choice was observed in the data. Since the QRE predictions depend on the logit precision λ, the likelihood function is evaluated for each λ to determine that it is maximized at $\lambda = 0.70$.[3]

The maximum likelihood estimate of λ was used to construct the quantal response curves shown earlier in Figures 2 and 4 for the paired treatment changes. In each case, the fitted QRE choice predictions are at the intersections of the decreasing defense quantal responses with the solid and dotted attack (increasing) quantal responses. In both cases, the QRE intersections move to the right, correctly predicting the observed increase in the α_1 attack rate after the treatment change documented in the previous section. These QRE predictions contrast sharply with the Nash attack rate predictions (of no change for equal absolute increases in *A* and *B*, and of a reduction for equal proportional increases).

Figure 7. Effects of Equal Absolute and Equal Proportional Increases in A and B



Note: The curved lines are quantal responses for the various treatments, with intersections at the squares, which represent QRE predictions. The dark triangles are data averages that are near the QRE predictions for each treatment.

Figure 7 shows a unified picture of QRE predictions (squares at intersections of quantal response lines) and dark data dots that indicate the data averages for the

[3] Since there is twice as much data for the *A*=6, *B*=12 treatments, we divided the data counts and number of observations for this treatment by 2 in the likelihood function to avoid over-weighting this treatment.

treatments (LowProp and High). The data dot in the lower right part shows the averages for the High treatment that was run twice, once paired with the LowAb treatment and once with the LowProp treatment. The QRE (like Nash) predictions are consistent with the sharp reduction in the d_2 frequency in the High treatment, but QRE correctly accounts for the most salient feature of the data, the increase in the a_1 attack frequency, which is the opposite of the Nash prediction of a decrease (from LowProp). The QRE intuition is that the enhanced effectiveness of the a_1 attack causes the attacker's upward sloping quantal response curves to shift to the right, as shown by the upward sloping dotted line.

Table 3 presents a numerical comparison of the treatment data averages with the theoretical predictions. The data points in the top row, in bold, are listed as (α, δ) pairs for a_1 and d_2 choice rates. The middle and bottom rows show analogous QRE and Nash point predictions by treatment. For the equal absolute change treatments on the left side of the table, the fitted logit QRE predictions in the middle row imply that the a_1 frequency should increase from 0.36 in the LowAb treatment (A = 2, B = 8) to 0.59 in the HighAb treatment (A = 6, B = 12). The QRE model fits the patterns of the data better than Nash, both qualitatively and quantitatively. This strong positive treatment effect on the a_1 attack rate is captured by QRE and inconsistent with the Nash prediction of no effect. Furthermore, the QRE estimates of attack frequencies in these two treatments closely track the observed frequencies in the data (0.33 and 0.59).

Table 3. Data, Nash, and Logit QRE: (α, δ)

	Equal Absolute Change		**Equal Proportional Change**	
Treatment Condition:	LowAb	HighAb	LowProp	HighProp
Data Averages	(0.33, 0.41)	(0.59, 0.19)	(0.49, 0.66)	(0.63, 0.17)
Logit QRE for λ = 0.7	(0.36, 0.46)	(0.59, 0.17)	(0.41, 0.58)	(0.59,0.17)
Nash Equilibrium	(0.33, 0.56)	(0.33, 0.11)	(0.50, 0.67)	(0.33, 0.11)

Note: The top row shows data averages (attack rate α, defense rate δ) for each of the four treatments. These data averages are closely approximated by the quantal response equilibrium predictions in the second row. In contrast, the bottom row Nash attack predictions of 0.33 for the high attack effectiveness columns are much lower than the data averages of 0.59 and 0.63 shown at the top of the second and fourth columns.

The same observation holds for equal proportional increases in a_1 attack effectiveness (the LowProp to HighProp change in the right two columns of Table 4), where the estimated QRE predictions of a_1 attack frequencies, 0.41 and 0.59 (middle row), closely track the observed increase from 0.49 to 0.63 (top row), in stark contrast to the Nash prediction of a decline from 0.50 to 0.33 (bottom row).

Table 3 also displays the data, Nash predictions, and QRE estimates for δ, the d_2 defense frequency, shown as the second number in each pair. While Nash and QRE both make the same qualitative prediction that d_2 defense frequencies will fall in the HighAb and HighProp treatments than in the Low treatments, the observed declines in the d_2 defense data (top row) are smaller in magnitude than the declines that the Nash model predicts (bottom row), and this dampening is reflected in the QRE estimates (middle row).

B. Concealed Payoff Parameter Estimation

The method used above for logit QRE estimation and comparison with Nash equilibrium was only possible because, in a laboratory setting, the payoff parameters of the games are known, fixed numbers. This contrasts with structural estimation methods used in applied field settings. In the analysis of auction data, for example, the payoff parameters like prize values are unknown to the econometrician and need to be estimated.

In this subsection, we follow this traditional field-data approach by ignoring the fact that the payoff parameters in the experiment are known, and instead blinding ourselves to the two parameters that are common to all treatments (C=7 and D=4). Then we estimate those parameters for both QRE and Nash equilibrium models. Since the true values of the estimated parameters are known with certainty, a simple statistical test can determine whether the concealed parameters are accurately recovered by the estimation for each of the two models.

We use maximum likelihood to estimate C and D using the logit QRE model with λ fixed at 0.7 and compare this to the maximum likelihood estimates of these two parameters using the Nash mixed strategy equilibrium model. The QRE estimates are $\hat{C}_{QRE} = 7.1$ and $\widehat{D}_{QRE} = 3.7$ (third row of Table 5) and the Nash estimates are $\hat{C}_{Nash} =$

7.6 and $\widehat{D}_{Nash} = 3.1$ (bottom row). For each model, we conduct a likelihood ratio test with *C*=7 and *D*=4 as the null hypothesis to see if the models successfully recover the true experimentally controlled parameters.

The other parameters in the game, *A* and *B*, vary across the three treatments, and are treated as known parameters in the estimation. Both QRE and Nash models are identified because we have three different (*A*, *B*) pairs in our treatment, each of which generates different QRE and Nash equilibrium strategy profiles for all values of *C* and *D* that satisfy inequalities (i) and (ii) in section II.

Table 4. Concealed Parameter Estimation (a_1 Attack Rate, d_2 Defense Rate)

	LowAb	**High**	**LowProp**	**−logLd**
Attack and Defense Data:	(0.33, 0.41)	(0.61, 0.18)	(0.49, 0.66)	
QRE estimates:				
$\lambda = 0.7$, C = 7, D = 4	(0.36, 0.46)	(0.59, 0.17)	(0.41, 0.58)	1832.675
$\lambda = 0.7$, $\hat{C} = 7.1$, $\widehat{D} = 3.7$	(0.38, 0.47)	(0.60, 0.18)	(0.45, 0.59)	1830.374
Nash estimates:				
C = 7, D = 4	(0.33, 0.56)	(0.33, 0.11)	(0.50, 0.67)	1924.214
$\hat{C} = 7.6$, $\widehat{D} = 3.1$	(0.43, 0.53)	(0.43, 0.15)	(0.60, 0.61)	1889.242

Note: This table shows estimates of the C and D payoff parameters of the game, which were known by the experimenter but "concealed" in the estimation. The resulting estimates for the QRE specification ($\hat{C} = 7.1$, $\widehat{D} = 3.7$) in the left column are significantly closer to the known values of 7 and 4 than the corresponding estimates using the Nash equilibrium specification.

The choice frequencies implied by parameter estimates are shown in the middle columns of Table 4, and the last column shows the log likelihood values. We conduct a likelihood ratio test for both the QRE and the Nash models, with *C*=7 and *D*=4 as the null hypothesis, where the test statistic (twice the likelihood ratio) is chi-square distributed with two degrees of freedom. The QRE estimates of $\hat{C} = 7.1$ and $\widehat{D} = 3.7$ are not significantly different from the true values (7, 4) at the 5% significance level (χ^2=4.602, p>0.095). For the Nash model, we obtain an estimate of $\hat{C} = 7.6$, $\widehat{D} = 3.1$, which are significantly different from (7, 4) at the 1% level (χ^2=69.94, p<0.001). The bottom line is that the QRE model successfully recovers the concealed parameters, C and D, while the Nash model does not.[4]

The difference between the success of the two models in recovering the payoff parameters is also reflected in the difference between the constrained and unconstrained estimates of the mixed strategy attack and defend frequencies in the three treatments. The average difference in the estimate for the attack rate α is 0.10 for the Nash model and 0.02 for the QRE model, while the average difference in the defense rate δ estimate is 0.04 for the Nash model and 0.01 for the QRE model.

VI. Related Applied and Empirical Studies

Our binary attacker-defender study can provide insights that may enhance an understanding of behavior in more complex naturally occurring conflict settings. To illustrate this, we discuss two specific applications. The first is an empirical study of rebel attacks on oil infrastructure facilities in African countries (Blair et al. 2022), modeled as a simultaneous-move game between an attacker ("rebel") and defender ("government") that has similar strategic incentives as our game. There are $N \geq 2$ possible oil infrastructure targets, with location 1 being a high-value critical facility (e.g. a refinery) and the other locations arrayed along a pipeline supply chain. The rebels choose a location to attack and the government chooses a location to defend. If the attacked location is defended, the attack is blocked, resulting in no gain for the attacker and no loss for the defender. If an attacked non-critical location is undefended, the attack is successful, resulting in a loss of $-(v_G + k)$ for the government and a gain of v_R for the rebel group, where k represents a government repair cost, and v_G and v_R represent the direct monetary loss to the government and the rebel gain, respectively. The analogous government and rebel payoffs from a successful attack at the critical infrastructure

[4] Alternatively, one could estimate the QRE model with *three* free parameters, λ, C, and D, instead of fixing the precision at 0.7. In this case, the estimated precision is $\hat{\lambda}$ = 0.75 and the hidden parameter estimates are essentially unchanged ($\hat{C} = 7.2,\ \widehat{D} = 3.7$) from the estimates of $\hat{C}$ and $\widehat{D}$ in row 3 of the table.

location are $-\theta(v_G + k)$ and θv_R respectively. The parameter $\theta > 1$ scales how much greater is the value of the critical location to both sides.

The case of $N = 2$ locations reduces to a game similar to our 2x2 binary conflict game, with the payoffs shown in the table below, where α denotes the probability of attacking location 1 and δ denotes the probability of defending location 2.

Two-Location Point of Attack Game

(adapted from Blair et al. 2022)

	a_1 (α)	a_2 ($1-\alpha$)
d_1 ($1-\delta$)	0, 0	$-(v_G+k)$, v_R
d_2 (δ)	$-\theta(v_G+k)$, θv_R	0, 0

The mixed Nash equilibrium is $\delta = \alpha = \frac{1}{1+\theta}$. Since $\theta > 1$, the rebel group is *more* likely to attack at location 2, despite its lower value, whereas the government is more likely to defend the critical infrastructure at location 1.[5] As in our model, these equilibrium conditions have comparative statics implications about how attacker and defender behavior changes with the underlying payoffs. For example, an increase in the world price of oil increases the scale parameter θ, which is similar to a proportional increase in a_1 attack effectiveness in our model. This scale increase is predicted, counter to naïve intuition, to result in *fewer* attacks on critical location 1. On the other hand, an increase in repair costs, k, has no effect on equilibrium attack frequencies, parallel to the non-effect of absolute attack effectiveness changes in the Nash equilibrium of our model.

These and other qualitative predictions are derived by Blair et al. (2022) for the more general model with more than one non-critical pipeline location. The authors use geo-spatial data for insurgent attacks in selected African countries to analyze the pattern of attack rates. The resulting estimates conform to the qualitative prediction that rebels

[5] Thus the model predicts a misalignment between attack and defense "profiles." See also Holt, et al. (2016) for a study of "misaligned profiling" in an airport security context, where defenders tend to target potentially effective attackers (young, single, athletic, international), whereas the attackers often rely on those who do not fit those profiles.

tend to focus their attacks on the less critical pipelines, and rebel groups randomize where they strike.

While those findings are in line with the game theoretic models, the reduced-form methodological approach differs from the structural approach taken in our paper. This limits the ability to directly test with their methods whether their data on infrastructure attacks is better explained by Nash equilibrium or quantal response equilibrium. Ideally, in order to make such a comparison one would have to structurally estimate the driving payoff parameters of the underlying game in a similar way as we did in section V.B, albeit with unknown "true" payoff parameters. But this seems impractical or very difficult to do without additional strong assumptions (or more granular data) about heterogeneity across terrorist groups and across targets, and such analysis would not have luxury of comparing estimated payoff parameters with known "true" payoff parameters.

This application illustrates in a couple of ways how laboratory experiments have a useful complementary role to empirical field studies. The most obvious is the ability to directly control (and thereby observe without error) variables that are difficult or impossible to measure in the field, and to reduce or eliminate confounding factors such as heterogeneity across targets or insurgent groups. This greater control in the laboratory makes the application of structural estimation methods to the data for the evaluation of alternative theoretical models a straightforward exercise, in contrast the challenges of a structural approach to the Blair et al. (2022) data. A controlled experiment also has the advantage of being able to precisely manipulate the variables in the model to test comparative static predictions, for example by altering factors like a safe escape route that only changes the attacker's payoffs, and hence, that would only affect *defender* choice probabilities in a mixed-strategy Nash equilibrium.

A second illustration of how laboratory experiments can complement applied studies is the design of defense strategies. The focus here is on defending against terrorist attacks or smuggling incursions at a major airport with multiple entrances that can be reinforced with mobile patrols (Yang et al., 2013). The authors consider the problem of designing an optimal strategy for positioning 3 defense assets at a subset of the 8 gates of the Los Angeles International Airport (LAX). The objective is to position the assets to minimize the expected number of successful incursions. The defender (TSA) is modeled as a perfectly rational actor. But recognizing that human attackers may not be perfectly rational, the authors use a logit precision (λ = 0.76) taken from a previously published study to construct a predicted quantal response attack strategy at LAX that is *imperfectly responsive* to differences in attacker expected payoffs for each gate target.

The authors evaluated the effectiveness of the BRQR defense strategy (defender best response to attacker quantal response) and compared it to other strategies based on different behavioral assumptions about the attacker, using 40 financially-motivated human subjects who played the attacker role in an Amazon Turk experiment. In most cases, the quantal-response-based (BRQR) defense was the most effective of these behavioral models in terms of preventing successful attacks by the human subjects (Yang et al., 2013, Figure 4). This application shows how quantal response equilibrium theory combined with laboratory experiments can be usefully applied to design optimal defense strategies in security applications.

VII. Conclusions

The experiment reported here was motivated in three ways. First, the games we consider capture some of the key strategic elements of bilateral conflict, a classic and important problem that arises in several areas of political science. Second, the workhorse theoretical paradigm for studying these models, the Nash equilibrium, provides unintuitive predictions about how behavior responds to changes in underlying payoffs in the game. Third, quantal response equilibrium, which modifies the Nash equilibrium by injecting payoff-responsive errors into choice behavior, yields more intuitive predictions about the effects of such changes. In particular, the effect of this behavioral noise is that defenders respond only partially to the increased effectiveness of an attack method. This partial responsiveness for defenders tends to reinforce the incentive for attackers to increase use of the attack method with enhanced effectiveness, resulting in intuitive attack patterns that differ from Nash predictions.

The laboratory experiment implements two distinct changes in the effectiveness of one of the attacker's two strategies, using equal *absolute* or equal *proportional* increases in the effectiveness of that strategy. In the first case, Nash equilibrium predicts no change of attacker behavior. In the second case, Nash equilibrium predicts that the improved strategy is used *less often*. Both predictions are sharply contradicted by the observed increase in a_1 attack rates following a switch from low to high effectiveness for that action. In contrast, the observed data patterns are consistent with QRE.

One might wonder whether it would have been possible to explain any data pattern with this model, but this is not the case. The discipline that ensures the *empirical content* for QRE in our setup derives from the requirement that the quantal response function in equations (3) and (4) are smooth and payoff responsive, in the sense that action choice probabilities are ordered by equilibrium expected payoffs. Better actions are chosen more often that worse actions. The results are nonparametric: they do not depend on the functional form of the quantal response functions, only its ordinal properties.

The use of Nash equilibrium is the standard approach for the structural estimation of game-theoretic models using naturally occurring data that arise in applied microeconomics. Similarly, the Nash paradigm is used in political science for empirical analysis in a wide range of substantively important areas: legislative organization (Diermeier et al. 2003, Diermeier et al. 2005); judicial behavior (Iaryczower and Shum 2012); crisis bargaining in international conflicts (Crisman-Cox and Gibilisco 2018, Lewis and Schultz 2003, Signorino 1999, 2003); the causes of civil war (Gibilisco and Montero forthcoming); and committee decision making (Lopez-Moctezuma 2019). As some of those studies have shown, QRE is a feasible and useful alternative to Nash equilibrium as a theoretical foundation for structural modeling. Our experimental findings suggest that in some empirical applications, when facing a choice between using Nash equilibrium and QRE, the latter approach (if feasible) might be the better choice. The findings suggest a similar message for formal theorists who develop game-theoretic models to study problems of conflict, agency, legislative behavior, voting and other political science applications.

Propositions 1 and 2 illustrate the fact that QRE analysis can be a tractable and relatively straightforward alternative to Nash equilibrium.

As a relatively novel methodological contribution, this paper used concealed parameter recovery as a method of validation of the Nash and QRE models. This method is closely related to such structural empirical studies, and our analysis demonstrates how this technique can be applied to experimental data. Structural payoff parameters that are experimentally controlled and known to the experimenter and the subjects are treated as unobserved ("concealed"), are estimated ("recovered") from the observed choices, and are compared to the true values of those parameters. We find that QRE provides accurate parameter recovery, while estimation based on the Nash equilibrium model fails.

References

Ansolabehere, Stephen, James M. Snyder, Jr. and Charles Stewart, III (2001) "Candidate Positioning in U.S. House Elections," *American Journal of Political Science*, 45(January): 136-159.

Blair, Graeme, Darin Christensen, and Michael Gibilisco (2022) "The Point of Attack: Where and Why Does Oil Cause Armed Conflict in Africa," *Working Paper*, California Institute of Technology: Pasadena.

Brady, David W., Hahrie Han, and Jeremy C. Pope (2007) "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly*, 32(February): 79-105.

Chiappori, Pierre-Andre; Levitt, Steve and Groseclose, Timothy (2002) "Testing Mixed Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer," *American Economic Review*, 92(4), 1138-1151.

Crisman-Cox, Casey and Michael Gibilisco (2018) "Audience Costs and the Dynamics of War and Peace," *American Journal of Political Science*, 62(3): 566-580.

Diermeier, Daniel, Hulya Eraslan, and Antonio Merlo (2003) "A Structural Model of Government Formation," *Econometrica*, 71(January): 27-70.

Diermeier, Daniel, Michael Keane, and Antonio Merlo (2005) "A Political Economy Model of Congressional Careers," *American Economic Review*, 95(1): 347-373.

Fearon, James D. (1994) "Signaling versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model," *Journal of Conflict Resolution*, 38, June), 236-69.

Fey, Mark and K. W. Ramsay (2007) "Mutual Optimism and War," *American Journal of Political Science*, 51: 738-54.

Gerber, Elisabeth R. and Rebecca B. Morton (1998) "Primary Election Systems and Representation," *Journal of Law, Economics, and Organization*, 14(October): 304-24.

Gibilisco, Michael and Sergio Montero (forthcoming), "Do Major-Power Interventions Encourage the Onset of Civil Conflict? A Structural Analysis," *Journal of Politics*.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey (2005) "Regular Quantal Response Equilibrium," *Experimental Economics*, 8: 347-367.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey (2016) *Quantal Response Equilibrium: A Stochastic Theory of Games,* Princeton, NJ: Princeton University Press.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey (2020) "Stochastic Game Theory for Social Science: A Primer on Quantal Response Equilibrium," in M. Capra, R. Croson, T. Rosenblatt, and M. Rigdon eds. *The Handbook of Experimental Game Theory*, Edward Elgar: Cheltenham, pp. 8-47.

Holt, Charles A., Andrew Kydd, Laura Razzolini, and Roman Sheremeta (2016) "The Paradox of Misaligned Profiling: Theory and Experimental Evidence," *Journal of Conflict Resolution*, 60: 482-500.

Holt, C.A. and A.E. Roth (2004) "The Nash Equilibrium, A Perspective," *Proceedings of the National Academy of Sciences of the USA*," 101(12), 3999-4002.

Iaryczower, Matias and Matthew Shum (2012) "The Value of Information in the Court. Get it Right, Keep it Tight," *American Economic Review*, 102 (February): 202-237.

Lewis, Jeffrey B. and Kenneth A. Schultz (2003) "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information," *Political Analysis*, 11(4): 345-365.

Lopez-Moctezuma, Gabriel (2019) "Sequential Deliberation in Collective Decision Making: The Case of the FOMC," *Working Paper*, California Institute of Technology.

McKelvey, Richard D. and Thomas R. Palfrey (1995) "Quantal Response Equilibria in Normal Form Games," *Games and Economic Behavior*, 10: 6-38.

Merlo, Antonio and Thomas R. Palfrey (2018), "External Validation of Voter Turnout Models by Concealed Parameter Recovery," *Public Choice*, 176 (1-2): 297-314.

Powell, Robert (2007a) "Defending Against Terrorist Attacks with Limited Resources," *American Political Science Review*, 101(3): 527-541.

Powell, Robert (2007b) "Allocating Defensive Resources with Private Information about Vulnerability," *American Political Science Review*, 101(4): 799-809.

Ramsay, Kristopher W. (2017) "Information, Uncertainty, and War," *Annual Reviews of Political Science*, 20: 505-527.

Roberson, Brian (2006) "The Colonel Blotto Game," *Economic Theory*, 29: 1-24.

Signorino, Curt (1999) "Strategic Interaction and the Statistical Analysis of International Conflict," *American Political Science Review*, 93(2): 279-98.

Signorino, Curt (2003) "Structure and Uncertainty in Discrete Choice Models," *Political Analysis*, 11(4):316-344.

Sandler, Todd and Harvey E. Lapan (1988) "The Calculus of Dissent: An Analysis of Terrorists' Choice of Targets," *Synthese*, 76:245-261.

Walker, Mark, and John Wooders (2001) "Minimax Play at Wimbledon," *American Economic Review*, *91*(5): 1521-1538.

Yang, Rong, Christopher Kiekintveld, Fernando Ordonez, Milind Tambe, and Richard John (2013) "Improving Resource Allocation Strategy Against Human Adversaries in Security Games," *Artificial Intelligence Journal*, 195: 440-469.

SUPPORTING INFORMATION FOR ONLINE PUBLICATION

TABLE OF CONTENTS

A. Proofs of Propositions

Proposition 1. *Under assumptions* (i), (ii), and (iii), *if the effectiveness parameters A and B increase by equal absolute amounts* to $A + \gamma$ and $B + \gamma$, *then the Nash equilibrium probability of attack* a_1 *is unchanged, but the QRE probability* α *of this attack method is increased. The QRE probability of defending against this attack* $(1-\delta)$ *increases. This result holds for any quantal response function, Q and for any positive precision,* λ.

Proof. The comparative statics prediction for the effect of equal absolute shifts in *A* and *B* by the same amount γ can be derived by totally differentiating (3) and (4) with respect to γ under the condition that $dA/d\gamma = dB/d\gamma = 1$, so there is no effect on the $(B-A)$ terms in these equations. The result of this differentiation is a pair of equations in $d\alpha/d\gamma$ and $d\delta/d\gamma$:

(A1) $$\frac{d\alpha}{d\gamma} = \lambda Q'(\cdot)\frac{dA}{d\gamma} + \lambda Q'(\cdot)[C-D+B-A]\frac{d\delta}{d\gamma} \quad \text{and}$$

$$\frac{d\delta}{d\gamma} = -\lambda Q'(\cdot)[C-D+B-A]\frac{d\alpha}{d\gamma}$$

Then substituting the $d\delta/d\gamma$ expression on the right into the first equation in (A1) and using the fact that $dA/d\gamma = 1$, one obtains:

(A2) $$\frac{d\alpha}{d\gamma} = \frac{\lambda Q'(\cdot)}{1+Q'(\cdot)Q'(\cdot)\lambda^2[C-D+B-A]^2} > 0 \ .$$

The ratio in (A2) is positive since the $Q'(\cdot)$ derivatives are positive, even though the arguments of these functions (not shown) are different. This implies that an equal absolute increase in the *A* and *B* parameters will increase the QRE a_1 attack probability, and vice versa, for any quantal response function *Q*. The second equation in (A1) then implies that $\frac{d\delta}{d\gamma} < 0$, so the $(1-\delta)$ probability of defending against attack a_1 increases.■

Proposition 2. *Under assumptions* (i), (ii), and (iii), *if the* a_1 *attack effectiveness payoffs, A and B*, *increase by equal proportional amounts* (*to* πA *and* πB, *where* $\pi > 1$), *then for any quantal response function Q satisfying (iii), there exists a positive precision parameter* λ_0 *such that for* $0 < \lambda < \lambda_0$, *the* QRE a_1 *attack rate with high effectiveness is higher*, *while the analogous Nash attack rate is lower. Furthermore, the frequency of defending against* a_1 *increases in* π *for all values of* λ.

Proof. Replace *A* and *B* with πA and πB, respectively, in equations (3) and (4). Denoting $x = C-D+\pi(B-A)$, equations (3) and (4) reduce to:

$$\alpha = Q(\lambda[\pi A - C + x\delta]) \quad \text{and} \quad \delta = Q(\lambda[C - D - x\alpha])$$

Since the derivative of *x* with respect to π is $B-A$, the total derivative of the QRE conditions with respect to π are:

(A3) $$\frac{d\alpha}{d\pi} = \lambda Q'(\cdot)\left[A + \delta(B-A) + x\,\frac{d\delta}{d\pi}\right] \quad \text{and}$$

$$\frac{d\delta}{d\pi} = -\lambda Q'(\cdot)\left[\alpha(B - A) + x\frac{d\alpha}{d\pi}\right].$$

These 2 equations can be solved for the two comparative statics derivative, and hence:

(A4) $\frac{d\delta}{d\pi} = -\lambda Q'(\cdot)\ \frac{\alpha(B-A)+x\lambda Q'(\cdot)[A+\delta(B-A)]}{1+\lambda^2 Q'(\cdot)Q'(\cdot)x^2}$

Since $Q' > 0$, it follows that $\frac{d\delta}{d\pi} < 0$, as stated in the proposition. The only negative term on the righthand side of the equation for $\frac{d\alpha}{d\pi}$ in (A3) is $\frac{d\delta}{d\pi}$, which vanishes for small λ, and hence $\frac{d\alpha}{d\pi} > 0$. It is easily verified that the effect on the Nash equilibrium goes in the opposite direction. ■

B. Supplemental Data Appendix

Average Attack and Defense Rates for Equal Absolute Increases in Attack Effectiveness

Treatment: Session and Attacker ID	***A*=2, *B*=8** *a*1 attack rate	***A*=2, *B*=8** *d*2 defend rate	***A*=6. *B*=12** *a*1 attack rate	***A*=6, *B*=12** *d*2 defend rate	change in *d*2 rate	change in *d*2 rate ≥ -0.45
prqr2 ID7	0.2	0	0.7	0.25	0.25	1
prqr2 ID8	0.15	0.6	1	0	-0.6	0
prqr2 ID9	0.55	0.45	0.25	0.1	-0.35	1
prqr2 ID10	0.1	0.65	0.1	0.75	0.1	1
prqr2 ID11	0.3	0.2	0.45	0.3	0.1	1
prqr2 ID12	0	0	0.45	0.2	0.2	1
prqr4 ID7	0.45	0.5	0.6	0.3	-0.2	1
prqr4 ID8	0.35	0.55	0.95	0	-0.55	0
prqr4 ID9	0.4	0.45	0.4	0.15	-0.3	1
prqr4 ID10	0.55	0.6	0.95	0	-0.6	0
prqr4 ID11	0.55	0.4	0.3	0.25	-0.15	1
prqr4 ID12	0.45	0.35	0.6	0.3	-0.05	1
prqs1 ID6	0.15	0.4	1	0	-0.4	1
prqs1 ID7	0.4	0.45	0.65	0.1	-0.35	1
prqs1 ID8	0.35	0.3	0.6	0.15	-0.15	1
prqs1 ID9	0.45	0.3	0.6	0.3	0	1
prqs1 ID10	0.1	0.65	0.1	0.6	-0.05	1
prqs3 ID8	0.4	0.05	0.95	0	-0.05	1
prqs3 ID9	0.35	0.45	0.6	0.15	-0.3	1
prqs3 ID10	0.3	0.45	0.5	0.1	-0.35	1
prqs3 ID11	0.4	0.45	0.3	0.15	-0.3	1
prqs3 ID12	0.3	0.65	0.65	0.2	-0.45	0
prqs3 ID13	0.2	0.35	0.95	0	-0.35	1
prqs3 ID14	0.45	0.6	0.55	0.2	-0.4	1
Average	**0.3292**	**0.4104**	**0.5917**	**0.1896**		

Notes: Treatment parameters *A* and *B* refer to attacker payoffs in Table 1 when the attacker (column player) chooses decision *a*1 (left column). The low payoff treatment (*A*=2, *B*=8**)** occurred in the first 20 rounds in sessions *prqr2* and *prqr4*, and in the final 20 rounds in sessions *prqs1* and *prqs3*. The rows show 20-period average decision rates for each 20-period block for each attacker ID of each session show on the left. Pairings were fixed in all periods, so there is one defender matched with each attacker ID. The variable definitions are:
***a*1 attack rate** = proportion of *a*1 (left column) attack decisions for the attacker ID in that row
d2 defend rate = proportion of *d*2 (bottom row of Table 1) decisions for the matched defender
change in *d*2 rate >= -0.45 = Boolean variable, 1 if the *d*2 rate fell by less than the -0.45 Nash prediction

Average Attack and Defense Rates for Equal Proportional Increases in Attack Effectiveness

Treatment: Session and Attacker ID	A=3, B=6 a1 attack rate	A=3, B=6 d2 defend rate	A=6. B=12 a1 attack rate	A=6, B=12 d2 defend rate	change in d2 rate	change in d2 rate ≥ − 0.56
prqu2 ID7	0.4	0.5	0.75	0.3	-0.2	0
prqu2 ID8	0.1	0.85	1	0.1	-0.75	1
prqu2 ID9	0.75	0.85	0.65	0.35	-0.5	0
prqu2 ID10	0.5	0.8	0.35	0.05	-0.75	1
prqu2 ID11	0.55	0.8	0.5	0.4	-0.4	0
prqu2 ID12	0.2	0.75	0.7	0.3	-0.45	0
prqu4 ID7	0.5	0.55	0.65	0.2	-0.35	0
prqu4 ID8	0.45	0.55	0.4	0.15	-0.4	0
prqu4 ID9	0.45	0.85	0.5	0.2	-0.65	1
prqu4 ID10	0.45	0.45	0.3	0.2	-0.25	0
prqu4 ID11	0.15	0.8	0.7	0.1	-0.7	1
prqu4 ID12	0.55	0.45	0.6	0.25	-0.2	0
prqu5 ID7	0.5	0.7	0.9	0.15	-0.55	0
prqu5 ID8	0.4	0.55	0.6	0.4	-0.15	0
prqu5 ID9	0.65	0.65	0.7	0.1	-0.55	0
prqu5 ID10	0.75	0.7	0.65	0.05	-0.65	1
prqu5 ID11	0.6	0.65	0.4	0.2	-0.45	0
prqu5 ID12	0.4	0.75	0.75	0.15	-0.6	1
prqu6 ID7	0.55	0.85	1	0.05	-0.8	1
prqu6 ID8	0.55	0.45	0.85	0.2	-0.25	0
prqu6 ID9	0.65	0.45	0.55	0	-0.45	0
prqu6 ID10	0.65	0.6	0.6	0.1	-0.5	0
prqu6 ID11	0.55	0.75	0.2	0.1	-0.65	1
prqu6 ID12	0.45	0.55	0.7	0.05	-0.5	0
Average	**0.4896**	**0.6604**	**0.6250**	**0.1729**		

Notes: Treatment parameters *A* and *B* refer to attacker payoffs in Table 1 when the attacker (column player) chooses decision *a*1 (left column). The low payoff treatment (*A*=3, *B*=6**)** occurred in the first 20 rounds in sessions *prqu4* and *prqu6*, and in the final 20 rounds in sessions *prqu2* and *prqu5*. The rows show 20-period average decision rates for each 20-period block for each attacker ID of each session show on the left. Pairings were fixed in all periods, so there is one defender matched with each attacker ID. The variable definitions are:

***a*1 attack rate** = proportion of *a*1 (left column) attack decisions for the attacker ID in that row

d2 defend rate = proportion of *d*2 (bottom row of Table 1) decisions for the matched defender

change in *d*2 rate >= -0.56 = Boolean variable, 1 if the *d*2 rate fell by less than the -0.56 Nash prediction

C. Instructions Appendix

Rounds and Matchings: The experiment consists of a number of **rounds**. Note: You will be matched with the **same** person in all rounds.

Interdependence: Your earnings are determined by the decisions that you and the other person make.

Roles: In each pair of people, one person will be designated as the "row" player and the other will be the "column" player. You will be a **column player** (or) **row player** in all rounds.

Continue with Instructions

The column player will press either the **Left** or the **Right** button. The row player will choose **Top** or **Bottom**. These choices determine which part of the matrix is relevant (Top Left, Top Right, Bottom Left, Bottom Right). In each cell, the row player's payoff is shown in blue and the column player's payoff is shown in red.

Payoff Matrix (Row, Column)

	Left	**Right**
Top	**$0.80, $0.40**	**$0.40, $0.80**
Bottom	**$0.50, $0.70**	**$1.00, $0.20**

- If you are a row player, your decision buttons will be on the left side of the payoff table, and if you are a column player, your decision buttons will be above the table.

Continue with Instructions

- **Matchings:** Please remember that you will be matched with the **same** person in all rounds.
- **Earnings:** Your earnings are determined by the choices that you and the other person make in the round. You begin with a fixed payment of **$0**, and earnings will be added to this amount (losses, if the game has negative payoffs, will be subtracted). Your total earnings will be displayed in a cumulative earnings column on the page that follows.
- **Rounds:** There will be **20 rounds** in this part of the experiment, and you are matched with the same person in all rounds.