Coordination and Learning in Dynamic Global Games: Experimental Evidence*

Ernst Fehr† Olga Shurchkov‡

Current Version: March 2010
First Version: September 2006

Abstract

Coordination problems are ubiquitous in social and economic life. Political mass demonstrations, the successful organization of a high profile scientific workshop, the decision whether to join a speculative currency attack, investment in a risky venture, and capital flight from a particular country are all characterized by coordination problems. However, our understanding of the forces behind the coordination of heterogeneous actors remains limited. Here we exploit recent developments in the theory of dynamic global games to study experimentally how agents with heterogeneous information about the underlying fundamentals achieve coordination. We document that individuals' willingness to attack the status quo depends negatively on the private signals about the strength of the status quo and the costs of attacking, and positively on the arrival of new, more precise, information after a failed attack. Although these findings support the qualitative predictions of the model, we also find that subjects are considerably and persistently more aggressive in their attacks than the theory predicts because they have overly optimistic expectations about others' attacking behavior. These above-equilibrium expectations are partially self-fulfilling due to the strategic complementarity in agents' actions, which may explain the persistence of individuals' excess aggressiveness.

Keywords: Coordination, Learning, Dynamics, Global Games, Crises, Rationality, Excess Aggressiveness, Experimental Economics

JEL Classification: C7, C9, D8, F3

*The authors are indebted to George-Marios Angeletos, for his time, helpful comments, and support throughout this project. We would also like to thank Daron Acemoglu, Miriam Bruhn, James Costain, Florian Ederer, Muhamet Yildiz, and the seminar participants and organizers of the MIT macroeconomics workshop, MIT macroeconomics and theory field lunches, Wellesley College Calderwood seminar, the IESE Conference on Complementarities and Information, the 2007 SAET conference, the conference of the French Economic Association on Behavioral Economics and Experiments, the ESA World Meeting 2007, and the 22^{nd} World Congress of the EEA for valuable comments and discussion. Finally, Olga Shurchkov would like to acknowledge the hospitality of the Institute for Empirical Research in Economics where the experimental sessions were conducted.

†Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland. e-mail: efehr@iew.uzh.ch.

‡Department of Economics, Wellesley College, 106 Central St., Wellesley, MA, USA.
e-mail: oshurchk@wellesley.edu.

1 Introduction

Coordination amongst economic agents is an essential element in many macroeconomic events. The ability of individual participants to agree on a specific course of action such as an attack on a currency peg, a bank run, or a riot can determine the ultimate outcome for the economy as a whole and may change the course of a country's history.

Currency crises and capital flights are events involving coordination that have received particular attention in recent history. Within the affected country, a crisis can have an enormous negative impact on economic growth and can cause political change and turmoil.[1] Nevertheless, our ability to predict the outcomes of such events and our understanding of the reasons behind their onset and timing remain limited.

In particular, previous empirical work has failed to establish a clear relationship between the state of the macroeconomic fundamentals and the probability of a crisis (Glick and Rose 1998, Rose 2001, Meese and Rogoff 1983). This relationship is important because, if stronger fundamentals reduce the likelihood of a crisis, policies geared toward improving them would be advisable in times when the currency seems vulnerable to a potential attack. Moreover, the strength of the fundamentals determines the need for a potentially costly policy intervention such as an interest-rate hike. However, once again, we lack clear empirical evidence on the impact of policy interventions on the occurrence of currency crises (Kraay 2003). Furthermore, while it seems reasonable that a failed speculative attack may signal that the fundamentals are not too weak, it is not clear whether costly defense policies actually promise a prolonged period of tranquility or whether the threat of a new attack is imminent.

This paper uses an experimental approach to address the aforementioned issues. Our starting point is a dynamic global game that captures the features of currency crises that seem to be essential for understanding these issues: (1) the coordination element of currency crises that arises due to strategic complementarities in agents' actions, (2) the heterogeneity of expectations about the underlying economic fundamentals among the agents, and (3) the fact that the agents' beliefs about their ability to induce a regime

[1] For example, in Indonesia, the Asian economic crisis of 1997-1998 caused the growth rate of real GDP to fall from 8.2 percent in 1996 to 1.9 percent in 1997 to -14.2 percent in 1998. The figures are similar in other affected countries (IMF, 2000). In all cases, these countries experienced rapid capital outflows and large exchange rate devaluations. Furthermore, in Indonesia, the crisis played a role in ending Suharto's long period of authoritarian rule when riots and demonstrations caused his political isolation, finally compelling him to resign. In Thailand and South Korea, democratic elections were held and opposition parties came to power for the first time since political liberalization (Freedman, 2004). The effects of these crises go beyond the borders of just one nation. Crises spread from the country of origin to other economies, threatening to cause worldwide contagion (Forbes, 2000, Boston, 2003).

change may vary over time. Unlike the common-knowledge models of crises that capture the first feature but abstract away from the second and the third features (Obstfeld 1996), the dynamic global game model delivers concrete testable predictions regarding the impact of fundamentals, policy, and information on the equilibrium outcomes. Thus, its predictions can shed light on the aforementioned issues.

While it would be desirable to test the predictions of the model using field data, several problems would arise with this approach. Field data contain additional forces not captured by the model that can limit our ability of testing the particular forces that the model focuses on. Furthermore, field data are riddled with endogeneity problems that are difficult to avoid since natural experiments are not readily available in the context of crises. For example, suppose we were to use cross-country panel data to test whether interest rate hikes have an effect on the probability of a currency crisis. For a proper test, we would need a source of exogenous variation in the interest rate (i.e., a natural experiment). However, countries that raise interest rates may do so because they are already in a crisis situation, which means that causality may go both ways. By contrast, in the laboratory, we can perfectly measure all the relevant payoff variables ("fundamentals") and can exogenously vary the relevant "policy variables."[2] Furthermore, one of the goals of this study is to understand the structure of individual agents' strategies and beliefs. However, field data on crises are typically at the aggregate level and contain no information about individual behavior and expectations. Thus, using field data would prevent us from investigating whether fundamentals, policies, and information affect outcomes and individual behavior also through expectations about the actions of others, as the theory predicts. By contrast, a laboratory experiment allows the experimenter to elicit subjects' beliefs.

Our experiment is based on a two-period variant of the dynamic global game developed

[2] Recently, several interesting studies have used field data to test the predictions of static global games. We view these studies as complementary to the experimental approach taken in this paper. Chen, Goldstein, and Jiang (2007) use field data on mutual fund returns to find evidence that is consistent with the idea that complementarities may generate financial fragility. Prati and Sbracia (2002) focus on the effects of uncertainty about the fundamentals on the propensity for speculative attacks using reduced form specifications motivated by static global games. Danielsson and Peñaranda (2007) use structural estimation with data on the yen-dollar exchange rates to examine the relationship between fundamentals and liquidity that derives from a static global game. These studies provide evidence consistent with different qualitative predictions of static global games but the data in these studies do not enable the researchers to examine the causal impact of the strength of the fundamentals, the arrival of new information, or the cost of attacking. Because of a lack of information about the prevailing field parameters it is also not possible to identify the existence and the driving forces of overly aggressive attacking behaviors (relative to the predictions of the theory). For this reason, we believe that experimental evidence can provide further useful insights about the strengths and weaknesses of the theory and the actual behavioral mechanisms behind coordination behavior.

by Angeletos, Hellwig, and Pavan (2007). The model consists of a large number of agents and two possible regimes, the status quo and an alternative to the status quo. The game continues into the second period as long as the status quo is in place. In each period, each agent can either attack the status quo (i.e., take an action that favors regime change), or not attack. The net payoff from attacking is positive if the status quo is abandoned in that period and negative otherwise. Regime change, in turn, occurs if and only if the percentage of agents attacking exceeds a threshold $\theta \in \mathbb{R}$ that parameterizes the strength of the status quo. The parameter θ captures the component of the payoff structure (the "fundamentals") that is never common knowledge, as is customary in the global games literature (Carlsson and van Damme 1993a, 1993b; Morris and Shin 1998). In the first period, each agent receives a private signal about θ. If the game continues into the second period, agents may or may not receive more private information about θ.

Within the context of currency crises, the fundamental θ represents the strength of the currency peg or the ability of the central bank to defend the peg. The agents are the speculators deciding whether to attack the currency. The cost of attacking can be interpreted as the interest rate. This framework has been applied to several macroeconomic phenomena: see Goldstein and Pauzner (2001) and Rochet and Vives (2004) for bank runs; Corsetti, Guimaraes and Roubini (2003) and Morris and Shin (2004) for debt crises; Atkeson (2000) for riots; Chamley (1999) for regime switches; and Edmond (2005) for political change.

The model makes the following predictions that are directly related to the aforementioned issues. In the first stage, the size of the attack is monotonically decreasing in the strength of the economic fundamentals, θ; equivalently, the agents' propensity to attack is decreasing in their individual private signals about θ. Secondly, an individual's propensity to attack, and, by implication, the aggregate size of the attack, decrease in the cost of attacking and increase in the individual's expectation about the size of the attack. In the second stage, if the agents do not receive an additional more precise private signal, then no agent attacks the status quo again. This result arises from the fact that agents have learned that the regime survived a past attack, which along the equilibrium path means that the fundamental θ must be good enough. We henceforth refer to this type of learning as "endogenous learning." Endogenous learning also reduces the agent's beliefs about the aggressiveness of others. On the other hand, under some parameter restrictions, a new attack becomes possible in the second stage if agents receive sufficiently precise new information. We henceforth refer to the arrival of new information as "exogenous learning." While "endogenous learning" reduces the incentive to attack in the second period, "exogenous learning" can make a new attack possible.

To address the policy issues discussed above, we conduct several treatments of a laboratory experiment where we vary the strength of the fundamentals, the cost of attacking, and the availability of information in the second stage.

In the first stage of the experiment, we find that the size of the attack is monotonically decreasing in both, the strength of the fundamentals and the cost of attacking, and that subjects' strategies are monotonic in their private signals which is consistent with the theory prediction of a unique equilibrium in monotone strategies. Furthermore, we find that subjects take a truly game-theoretic approach to deciding whether or not to attack the status quo. This means that the relative strength of the status quo and the cost of attacking impact observed behavior mainly through agents' beliefs about the actions of others. This type of "strategic reasoning" is consistent with the theory.

In the second stage of the experiment, we examine the effects of the interaction between learning and coordination. In order to distinguish between the effects of endogenous and exogenous learning, we first run a treatment where the subjects do not receive any new private information in the second stage. In this case, the only additional information that the subjects receive in stage two is that the game has not ended. In equilibrium, this means that the fundamentals are sufficiently good. Rational subjects should therefore infer this from the fact that the status quo has survived. We find that, in the second stage, the subjects are in fact learning from the outcome of the first stage since the probability of attack is greatly reduced.

Next, to examine the impact of exogenous learning, we run a treatment where the subjects receive an additional precise private signal in the second stage of the experiment. In this case, the subjects are still able to learn endogenously through their observation that the experiment proceeded into stage two, but in addition, they can now learn exogenously by incorporating this more precise information into their decision to attack the status quo. We find that the probability of attack in the second stage now increases significantly relative to the treatment with endogenous learning only. Together, the second-stage findings imply that a policy-maker, having previously successfully defended the regime, cannot be assured that the crisis is averted. The endogenous learning, induced by the observation of a failed attack, alone makes the speculators relatively less aggressive, but a new attack may become possible as the agents accumulate new information about the strength of the regime.

While the evidence supports the qualitative predictions of the theory, we also find that the subjects' actions are overly aggressive relative to the model's predictions, at least in the treatments with a relatively high cost of attacking. Furthermore, we find that the agents' behavior is less responsive to the changes in the cost of attacking than in the theory. The policy implication is that interventions that target the cost of a speculative attack through interest rate hikes, for instance, may not be as effective as the theory suggests.

In order to explore the source of subjects' excess aggressiveness, we test the following two hypotheses. The first hypothesis postulates that the subjects are "irrational" in the sense that they are intrinsically biased toward attacking. An alternative hypothesis is that the agents act rationally given their beliefs of what others will do, and it is their beliefs that are more aggressive than the theory predicts. We find evidence that leads us to reject the first hypothesis: given the subjects' aggressive expectations relative to the model predictions, their actions are mostly consistent with best-response strategies. The subjects' aggressive behavior stems from their aggressive beliefs and not from intrinsic aggressiveness. Furthermore, the persistence of the aggressiveness of beliefs over time is partly justified by the self-fulfilling nature of expectations. Subjects not only believe that the status quo is more likely to be successfully overturned than the theory predicts, but they also experience this to be the case.[3]

This paper is related to the theoretical and experimental literature on coordination games with common knowledge in a static environment (Cooper et al. 1990, 1992; Van Huyck, Battalio, and Beil 1990) and in the dynamic environment (Cheung and Friedman 2006; Brunnermeier and Morgan 2006).[4] Experimental papers that are most closely related to the present study involve laboratory tests of the predictions of static coordination games with private information. Cabrales, Nagel, and Armenter (2003) test the *global* coordination game theory in two-person games with random matching inspired by Carlsson and van Damme (1993a), while Heinemann, Nagel, and Ockenfels (2004) test the predictions of the static speculative-attack model of Morris and Shin (1998). To our knowledge, this study is the first to test the predictions of *dynamic global* games in a laboratory experiment. We are also not aware of any other work that explores the extent

[3]Our results therefore also relate to the literature on the role of persistent deviations from equilibrium play in environments characterized by strategic complementarity (Haltiwanger and Waldmann 1985, 1989; Fehr and Tyran 2005, 2008). This literature emphasizes the fact that under strategic complementarity the rational agents have an incentive to partially mimic the irrational agents which reinforces deviations from equilibrium play at the aggregate level. In our setting the mere belief that other players are deviating from the equilibrium frequency of attacking the status quo induces rational players to increase the probability of attacking, which partially reinforces the initial beliefs. This reinforcement may make it difficult to converge to the equilibrium. See also Izmalkov and Yildiz (2008) for a general class of models where agents' sentiments are believed to play an important role. Arifovic and Masson (2007) build and simulate a model of a currency crisis where agents' beliefs are the source of volatility that can lead to a devaluation.

[4]Other recent studies examining effects of dynamics and coordination include Costain, Heinemann, and Ockenfels (2007) and Schotter and Yorulmazer (2003).

of agents' ability to use strategic reasoning and the role of self-fulfilling expectations in a dynamic coordination environment. In particular, our study appears to be the first to discover persistent excess aggressiveness in attacking behavior driven by subjects' excessively aggressive expectations about others' actions and to document that these expectations are partially self-fulfilling due to the presence of strategic complementaries in this setting.

The rest of the paper is organized as follows. Section 2 introduces the model and derives the theoretical predictions to be tested. Section 3 describes the experimental procedures and treatments. Section 4 discusses the variables used in the data analysis. Section 5 describes the experimental results. Section 6 concludes and discusses potential implications of the results.

2 The Model

2.1 The Environment

Our model is a two-period version of the model developed by Angeletos, Hellwig, and Pavan (2007). There are two regimes, the status quo and the alternative to the status quo.

Actions, Outcomes, and Payoffs. There is a continuum of agents, indexed by $i \in [0, 1]$, simultaneously deciding between two possible courses of action. Agent i can either choose action A ("attack"), an action that favors regime change, or choose action B ("not attack"), an action that favors the status quo. An "attack" can be interpreted as a speculative run on a currency, large withdrawal of funds from the economy's financial sector, or a political uprising against the government.

We denote the regime outcome by $R_{t+1} \in \{0, 1\}$ where $R_{t+1} = 0$ refers to the survival of the status quo, while $R_{t+1} = 1$ refers to the collapse of the status quo. Similarly, we denote the action of an agent by $a_{it} \in \{0, 1\}$, where $a_{it} = 0$ represents action B ("not attack"), while $a_{it} = 1$ represents action A ("attack").

Action A is associated with an opportunity cost c. If action A is successful (i.e., $R_{t+1} = 1$), each agent choosing action A earns an income of $y > c$. If not (i.e., $R_{t+1} = 0$), then the agent choosing action A earns $0 < c$. Action B yields no payoff and has no cost. Hence, the utility of agent i is

$$u_{it} = a_{it}(yR_{t+1} - c).$$

Finally, the status quo is abandoned ($R_{t+1} = 1$) if and only if

$$A_t \geq \theta$$

where $A_t \equiv \int a_i d_i \in [0, 1]$ denotes the mass of agents attacking at time t and θ parametrizes the exogenous strength of the status quo (or the quality of the economic fundamentals). A low value of θ represents a relatively weak state of the fundamentals, and a high value of θ represents a relatively strong state of the fundamentals.

If the regime change occurred ($R_{t+1} = 1$), there are no further actions to be taken in period 2. However, if the status quo survived ($R_{t+1} = 1$), the agents again have the opportunity to attack the status quo in period 2.

Complementarity. Note that the actions of the agents are strategic complements, since it pays for an individual to attack if and only if the status quo collapses and, in turn, the status quo collapses if and only if a sufficiently large fraction of the agents attacks. This strategic complementarity can be seen more clearly if we rewrite the payoff of an individual agent as

$$u_{it} = U(a_{it}, A_t, \theta) = \begin{cases} a_{it}(y - c) \text{ if } A_t \geq \theta \\ -a_{it}c \text{ if } A_t < \theta \end{cases}.$$

Consider for a moment the case where the state of the fundamentals, θ, is common knowledge among the agents. In that case, each agent's best response function is

$$g(A_t, \theta) = \underset{a_{it}\in\{0,1\}}{\arg\max} U(a_{it}, A_t, \theta) = \begin{cases} 1 \text{ if } A_t \geq \theta \\ 0 \text{ if } A_t < \theta \end{cases}.$$

Note that both $U(\cdot)$ and $g(\cdot)$ are increasing in A_t. It follows that, for all $\theta \in [0, 1]$, there are two pure-strategy Nash equilibria in this game, namely that either all agents choose action A or all agents choose action B; $[0, 1]$ is the region where multiplicity is possible. This is the type of multiplicity underlying all models of crises with strategic complementarities (such as Obstfeld 1996).

Information. However, in our setup (as in any global game), agents have heterogeneous information about the strength of the status quo. Nature draws θ from a normal distribution $N(z, 1/\alpha)$ which defines the initial common prior about θ.[5] Each agent then

[5] Note that z can also be thought of as a public signal that all agents receive. In particular, in the case when the prior is uniform, z represents a public signal and the rest of the model remains identical to the one we present here.

receives a private signal $x_{it} = \theta + \frac{1}{\sqrt{\beta_t}}\varepsilon_{it}$, where $\varepsilon_{it} \sim N(0,1)$ is i.i.d. across agents and independent of θ and β_t is the precision of private information.[6]

2.2 First-Period Predictions

In this section we present the predictions of the model for the first period. In the appendix we derive these predictions formally but here we focus primarily on the economic intuitions behind the results. Intuitively, if the status quo is very weak ($\theta < 0$), the regime collapses even if no one attacks and $A = 0$, because the zero attack is still greater than the negative fundamental. In that case, all individuals will receive low private signals x_{1i} which will induce all of them to attack the status quo.[7] Likewise, if the status quo is very strong ($\theta > 1$), the regime will survive even if everyone attacks and $A = 1$, because the fundamental will still be greater than 1. In that case, all individuals will receive high private signals which will induce them to refrain from attacking. More formally, it is strictly dominant to attack for sufficiently low signals, namely for $x_1 < \underline{x}$, where $\underline{x}$ solves $E[U(a_1, A(\theta), \theta|\underline{x}] = a_1(yPr[\theta \leq 0|\underline{x}] - c) = 0$ or after rearranging $Pr[\theta \leq 0|\underline{x}] = c/y$.[8] Similarly, it is strictly dominant not to attack for sufficiently high signals, namely for $x_1 > \overline{x}$, where $\overline{x}$ solves $Pr(\theta \leq 1|\overline{x}) = c/y$.

The previous argument suggests that for intermediate strengths' of the status quo some individuals receive signals that will induce them to attack while others receive signals that render it optimal for them to refrain from attacking. Moreover, in this intermediate range of θ's, an increase in the strength of the status quo will tend to increase the share of individuals with high private signals, implying that a higher share of individuals will refrain from attacking. Thus, in this intermediate range, a rise in θ will reduce the share of agents attacking the status quo, $A(\theta)$.

We depict the share of agents attacking the status quo, $A(\theta)$, in Figure 1. This figure also provides a neat way to graphically represent the economic forces behind the equilibrium.

In Appendix A, we show formally that it is optimal for an agent to follow a threshold

[6] The information structure is parameterized by $\beta_t = \sigma_{x,t}^{-2}$ and $\alpha = \sigma_z^{-2}$, the precisions of private and public information, respectively, or equivalently by the standard deviations, $\sigma_{x,t}$ and σ_z. Thus, $\alpha + \beta_t$ is the overall precision of information. The agents know the values of z, α, and β_t.

[7] Hereafter, we supress the subscript i for notational tractability.

[8] $E[U(a_1, A(\theta), \theta|\underline{x}]$ represents the expected payoff from attacking ($a_1 = 1$) given the private signal $\underline{x}$. Receiving signal $\underline{x}$ makes the agent indifferent between attacking and not attacking, and any signal below $\underline{x}$ guarantees that $\theta < 0$ which means that the regime collapses no matter what. $Pr[\theta \leq 0|\underline{x}]$ represents the posterior probability of $\theta \leq 0$ (regime collapse) given signal $\underline{x}$. In deriving $\underline{x}$, we use the definition of the utility function above.

strategy, that is, to attack the status quo if and only if he receives a private signal x_i below a threshold value x_1^*. As a result, for low enough realizations of θ, sufficiently many agents receive private signals below the threshold such that the status quo is abandoned (i.e., $A(\theta) > \theta$). Likewise, for high enough realizations of θ, sufficiently many agents receive private signals above the threshold such that the status quo is not abandoned (i.e., $A(\theta) < \theta$). It follows that there is a first-period equilibrium threshold θ_1^* such that for realizations of θ below θ_1^* the attack will be successful whereas for $\theta > \theta_1^*$ the attack will not succeed (θ_1^* is represented graphically as the point of intersection of the 45-degree line with $A(\theta)$ in Figure 1).

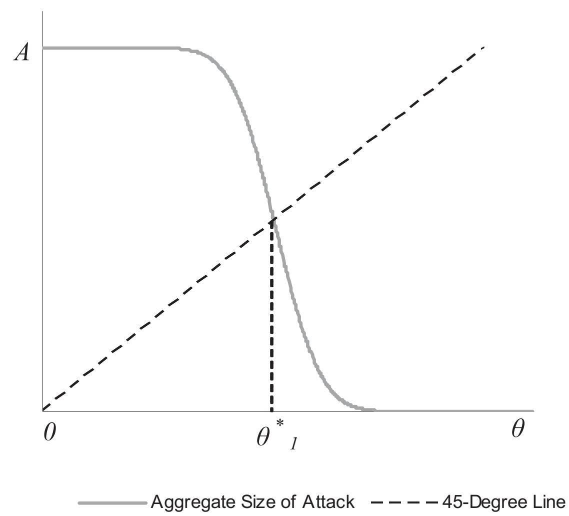


Figure 1: Aggregate Size of the Attack, A, for Different Realizations of θ, and the Equilibrium Threshold θ_1^*.

Figure 1 is also useful for illustrating an agent's view of aggregate attacking behavior if his beliefs about others' aggressiveness are overly optimistic or pessimistic. Figure 1 is based on the assumption that all agents know that all others follow a threshold strategy described above (attack the status quo if and only if they receive private signals below the critical value x^*)[9]. Suppose, however, that an agent erroneously believes that the others' threshold is $\widehat{x} > x^*$. This agent will then overestimate the share of people attacking for given realizations of θ, i.e., the agent will assume that others are overly

[9]For now, we also supress the subscript $t = 1$ for notational tractability.

aggressive. In graphical terms, the increase in the aggressiveness of agents' beliefs can be represented by a rightward shift of the function $A(\theta;\widehat{x})$ (see Figure 2). Moreover, because of the complementarity in the agents' attacking behavior this agent will now herself be more willing to attack. Thus, overly aggressive beliefs about others' behavior will lead to overly aggressive individual attacking behavior.

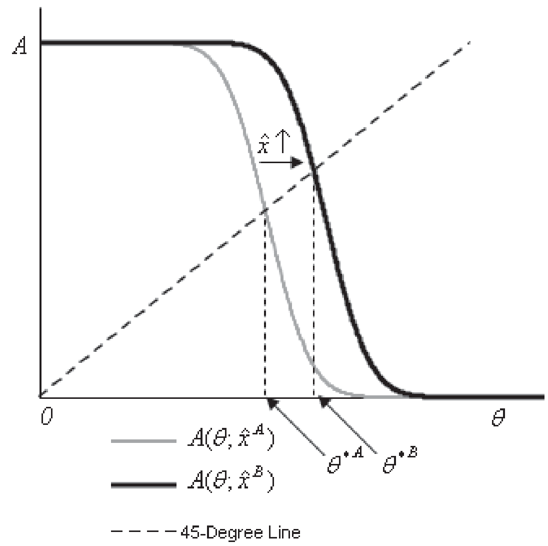


Figure 2: The Aggregate Size of Attack
for Different Realizations of θ and
Comparative Statics on $\widehat{x}$ ($\widehat{x}_A < \widehat{x}_B$).

A decrease in the cost of attacking is qualitatively equivalent to an increase in the perceived aggressiveness of others and leads to a larger share of agents attacking the status quo for all realizations of θ (represented by the shift to the right of the $A(\theta)$ curve in Figure 2). In terms of payoffs, the decrease in the cost or an increase in perceived aggressiveness of others makes it more attractive for the marginal agent to attack the status quo. In equilibrium, the threshold θ^* increases and the status quo collapses for a larger range of fundamentals.

The following are the first-period predictions of the model that we are able to test experimentally.

Prediction 1. **(Stage 1 Behavior)**

(a) The individual probability of attack is decreasing in the private signal, x, and equivalently, the size of the attack, $A(\theta;\widehat{x})$, is monotonically decreasing in θ.

(b) The individual probability of attack and the aggregate size of the attack are also decreasing in cost of attacking, c.

Prediction 2. **(Stage 1 Beliefs)** *The individual probability of attack is increasing in the reported belief about the aggregate size of the attack. This belief is decreasing in the private signal, x.*

2.3 Second-Period Predictions

Recall that the game continues into the second period if and only if the status quo remains in place in the first stage. This section present the predictions for the continuation game in the second period. Though the agent does not observe the first-period size of the attack, she carries over her first-period beliefs about the aggressiveness of others into the second period. In addition, she forms a new belief about the aggressiveness of others in the second period which need not be the same as the first-period perceived aggressiveness level.

In addition, the agents may or may not receive new private information in the second period. First, suppose that the agents receive no additional private signal in period 2. In this case, when agents arrive at the second period, they observe that the status quo has survived the first-period attack. From this observation, the agents can infer that the state of the fundamentals is not too weak, because otherwise the status quo would have collapsed under the first attack. In particular, it is now common knowledge that θ is above the first-period threshold θ^*. This knowledge causes agents' beliefs about others attacking behavior to become less aggressive, which makes agents' own behavior to become less aggressive, as well. This effect, in turn, guarantees that no agent is willing to attack in the second period. We refer to the effect on behavior of the continuation of the regime into the second period without new information as endogenous learning.

In theory, since every agent should know that others will not be attacking the status quo, the belief about the aggregate size of attack in this case should be reduced to zero as well. However, note that off equilibrium agents may form a new individual idea about the aggressiveness of others, $\widehat{x}_2$, which may cause them to become more aggressive even in absence of new information.

Next, we explore what happens if we change the information structure in the second period, such that agents receive an additional private signal that is sufficiently precise. In particular, what happens if the agents receive a signal, $x_{i2} = \theta + \xi_{i2}$, where $\xi_{i2} \sim N(0, 1/\beta_2)$

and β_2 is sufficiently high? We refer to the effects of the arrival of new information in the second period as exogenous learning.

With new information, the expected payoff from attacking, and therefore the individual probability of attacking, decrease in the private signal, x_2, and increase in the second-period aggressiveness of other agents, $\widehat{x}_2$. This prediction is equivalent to the first-period result and differs from it only insofar as the posterior about θ has been truncated for all the realizations of θ below θ_1^*. The novel feature of the continuation game is that the payoff from attacking in the second period is decreasing in the first-period aggressiveness of other agents, $\widehat{x}_1$. To better understand this new proposition, suppose that an agent believed that others attacked in the first period for a greater range of private signals, i.e., for this agent $\widehat{x}_1 > x_1^*$. This agent will therefore have a higher belief about the strength of the fundamentals (compared to an agent who believed that others applied threshold x_1^*), since in her mind the regime must have been strong enough to have survived a larger first-period attack. Thus, for given values of x_2 and $\widehat{x}_2$, this agent is strictly less likely to attack the status quo in the second period. On the other hand, the arrival of new information per se may cause the agents to become more aggressive. Which effect dominates depends on the prior, z; in particular, for a sufficiently high prior, z, the arrival of new information may result in the possibility of a new attack. The intuition comes from a combination of two effects. First, the arrival of new information causes the agents to discount endogenous learning. This effect alone makes them relatively more aggressive. However, the second effect of new information is that agents now also discount the prior, z. If z is low (an "aggressive prior"), discounting it makes the agents relatively less aggressive. If z is high ("lenient prior"), discounting it makes the agents relatively more aggressive and may eventually offset the incentive not to attack induced by the knowledge that the regime survived past attacks. As a result, there exists an x_2^* and an equilibrium in which an agent chooses to attack if and only if $x_2 < x_2^*$.

The reported belief about the size of the attack in the second period depends the two private signals, the observation of the survival of the status quo, and the individual perceived aggressiveness of other agents in the first and in the second period ($\widehat{x}_1$ and $\widehat{x}_2$). Similarly to the first period, the individual probability of attack is increasing in this belief, and in turn the belief is decreasing in the private signal.

The following are the second-period predictions of the model that we are able to test experimentally.

Prediction 3. **(Stage 2 Behavior, No New Info)** *If no new information arrives in the second period, no agent would attack the status quo.*[10]

Prediction 4. **(Stage 2 Beliefs, No New Info)** *If no new information arrives in the second period, agents' beliefs about the aggressiveness of others are lowered relative to the first period.*

Prediction 5. **(Stage 2 Behavior, New Info)** *If a new private signal arrives in the second period, for a sufficiently high prior, an attack becomes possible. In this case, the individual probability of attack is decreasing in the private signal,* x_2.

Prediction 6. **(Stage 2 Beliefs, New Info)** *If a new private signal arrives in the second period, the individual probability of attack is increasing in the reported belief about the aggregate size of the attack. This belief is decreasing in the private signal,* x_2.

In summary, we intend to test the following predictions in a laboratory experiment. In the first period, the aggregate size of attack is decreasing in both, the fundamental, θ, and the cost of attacking, c. The individual incentive to attack and the individual belief about the expected size of attack are decreasing in the private signal, x, and in the perceived aggressiveness of other agents. In the second period, a failed attack causes a reduction in the size of attack with no new information. However, the arrival of new information increases the incentive to attack.

3 Overview of the Experiment

3.1 Procedures

We conducted six sessions of the experiment at the experimental laboratory at the Institute for Empirical Research in Economics at the University of Zurich. The subjects were all students at the University of Zurich. The general procedures were kept the same throughout all six sessions. All sessions were computerized using the program z-Tree (Fischbacher, 2007). The subjects were first asked to read through and sign informed consent forms for non-biomedical research. Paper copies of the instructions were distributed to the participants prior to the beginning of the experiment. The subjects were given several examples clarifying the experimental procedures. Questions were answered in private. The subjects could not see or communicate with one another. At the end of the experiment, each participant filled out a computerized questionnaire. The questionnaire

[10] The robust prediction of the model is that the observation that the status quo has survived a first-period attack creates a large drop in the size of the attack in the second period. The prediction that the size of the attack drops exactly to zero is an artifact of the simple model. For example, this would not be the case if we were to allow θ to vary slightly across periods (see Angeletos, Hellwig, and Pavan (2007)).

asked the subjects about their strategies, as well as their understanding of statistics and probability.[11] At the very end, each subject was paid in cash a show-up fee equal to 15 Swiss Francs (CHF) and his or her earnings over the course of the session. Final income of each subject was first given in points and then converted to Swiss Francs, so that the average income (including the show-up fee) across all sessions was 51.6 CHF. Each session lasted approximately 1.5 hours.

Each of the six experimental sessions had 30 participants divided randomly into two groups of fifteen people. Each session consisted of 40 independent rounds of play, with each round corresponding to a new random number θ drawn from a normal distribution $N(z, 1/\alpha)$.[12] Thus, one can interpret each round as a new economy parametrized by the state of fundamentals, θ. The subjects were informed of the mean and the standard deviation of this distribution in the instructions. In addition, at the beginning of the round, each subject received a hint number (private signal, x) about the random number θ. In the instructions, the subjects were informed that the hint number was centered around the true value of θ and were given its standard deviation $(1/\beta_1)$.

Each round consisted of one or two stages (periods) of decision-making. In stage 1 of each round, each subject had to decide between actions A or B as described in section 2. Once all the subjects had chosen their actions in each stage of every round, they were asked a follow-up question: "How many other members of your group do you think chose action A?" Next, each subject received the following information: if the game ended after stage 1, he or she found out that action A was successful and learned the value of the unknown number θ, the number of other subjects choosing action A, and the payoff in that round. If the status quo survived such that the game continued into the second stage, two scenarios were possible. In the treatment without new information, the subject did not get an additional hint number (private signal) after the first stage, but only got a reminder of his or her first-stage hint number and received notification that action A was not successful. In the treatment with new information, the subjects received a new hint number if the game continued into stage 2. Analogously to the first stage, the subjects were informed that the second-stage hint number was centered around the true value of θ and were given its standard deviation $(1/\beta_2)$ in the instructions. At the end of the second stage, the subjects learned whether action A was successful, the value of the unknown number θ, how many other subjects chose action A in the first and in the second stage,

[11] Copies of the consent forms, intructions, and questionnaire questions in German or English are available upon request.

[12] To ensure the subjects' full understanding during the experiment, we provided them with several examples in order to familiarize them with the normal distribution.

and the payoff for the round.

3.2 Parameterization

We re-scaled all numbers by a factor of 100, so that the subjects did not have to deal with fractions. We chose the gross payoff, y, of a successful attack to be 100 and the gross payoff of an unsuccessful attack to be 0. This payoff scheme was chosen for its simplicity for the theory, as well as for the experiment participants.

Table 1 records the remaining parameters by session. In all cases, z, the prior about θ

Table 1: Parameters Chosen for the Experiment

Session	z, $1/\alpha$	$1/\beta_1$	$1/\beta_2$
1-4	65, 50	7	–
5-6	75, 55	10	1

was chosen to be high enough that a new attack becomes possible with the arrival of new information in the second stage. At the same time, in order to get a reasonable number of random draws within the critical interval of $[0, 100]$, we kept z sufficiently high and α sufficiently low. The standard deviation, β_t, was chosen based on satisfying the criterion for stage-one uniqueness of equilibrium, namely $\beta_1 \geq \frac{\alpha^2}{2\pi}$.

The parameters chosen for this experiment and recorded in Table 1 imply the theoretical thresholds (θ_1^*, x_1^*) given in Table 2.

Table 2: Predicted Thresholds

Cost	θ_1^*	x_1^*
20	81.5	87.8
50	48.1	47.8
60	34.8	30.9

We ran different treatment conditions based on the cost of action A and on the information provided to the participants in the second stage of the experiment. The various treatment conditions are summarized in Table 3.

In sessions 1 and 2, we began the experiment with the Cost 20 treatments, followed by the Cost 50 treatments. In order to test for any potential order effects, we reversed the cost treatment order in sessions 3 and 4. In all sessions 1-4, we maintained the same

Table 3: Treatment Summary.

	Cost of Action A		New Private Signal in Stage 2	
Session	First 20 Rounds	Second 20 Rounds	First 20 Rounds	Second 20 Rounds
1,2	**20**	**50**	No	No
3,4	**50**	**20**	No	No
5	60	60	**No**	**Yes**
6	60	60	**Yes**	**No**

information structure of the experiment. Similarly, in session 5, we began the experiment with rounds that did not provide the subjects with an additional private signal in stage 2, followed by rounds that provided this new information in stage 2. The order of the information treatments was reversed in session 6. In sessions 5-6, we maintained the same cost of attacking.

4 Variables and Summary Statistics

In our analysis, the main dependent variables are the aggregate size of the attack (measured as the fraction of subjects choosing action A), and the individual decision whether to attack (a binary choice variable, with 1 representing action A and 0 representing action B). We also consider the effects of the explanatory variables on the reported expectation of the number of other subjects choosing action A (*belief* variable).[13] This allows us to study the extent to which the observed outcomes are driven by strategic considerations. Henceforth, whenever we refer to the evidence of strategic reasoning in the data, we mean that actions are driven by the reported expectations of others' actions.

The main explanatory variables are the random number, θ, on the aggregate level and the subject-specific hint number (private signal, x) on the individual level. We also look at several other variables that can have important differential effects on the outcomes. One of these variables is the *cost* of action A (attacking), which varies across treatments and sessions and takes on values of 20, 50, and 60 (see Table 3). When we explore the impact of endogenous learning, we look at the effect of *stage* on actions, where stage takes on values of 1 or 2. In order to understand the effects of exogenous learning, we introduce a new-information dummy (*NI dummy*), which takes on a value of 1 in the treatment where subjects receive a more precise private signal in the second stage and a

[13]The belief variable takes on values 0-14.

value of 0 otherwise.

For robustness, we also include the following controls. The time trend for each session is introduced via the variable *round* which allows us to see whether subjects are learning about the experiment from round to round. This potential learning effect is explored further with the introduction of the dummy variable that takes on a value of 1 in the first ten rounds of a session and 0 otherwise, as well as the interactions between this dummy and our main explanatory variables: x, *cost*, *NI dummy,* and *belief.* To control for potential treatment order effects, we consider the *cost order* dummy and the *information order* dummy. The former takes on a value of 1 in sessions 1 and 2 where the Cost 20 treatment preceded the Cost 50 treatment and a value of 0 in sessions 3 and 4 where the Cost 50 treatment came first. The latter takes on a value of 1 in session 5 where the no-new-information treatment preceded the new-information treatment and a value of 0 in session 6 where the new information treatment came first.

Finally, taking advantage of the panel structure of the dataset, we run specifications controlling for subject and round *fixed effects*, clustering the standard errors at the group level.

Table B1 in Appendix B provides descriptive statistics for the experiment.

5 Results

In this section we present our data analysis and main results. In subsection 5.1, we focus on the effects of private information, the cost of action A, and expectations on behavior in the first stage. We begin by comparing actual behavior to the theoretical prediction. The systematic deviations from theory are explained in light of the subjects' expectations about others' actions. Subsection 5.2 concentrates on the effects of learning about the outcome of the first-stage attack as the subjects observe that the experiment continues into the second stage. Finally, in subsection 5.3, we investigate whether access to new, more precise private information in the second stage affects the subjects' actions and expectations.

5.1 Determinants of Attacks on the Status Quo

Our first result concerns the probability of attack of the status quo in the first stage of the experiment. Recall that the model in section 2 suggests that the probability of attack in the first stage is decreasing in the strength of the fundamental and in the cost of attacking. The main experimental findings can be summarized as follows:

RESULT 1 **(Stage 1 Behavior).** *(a) The individual probability of attacking the status quo (choosing action A) in stage 1 is decreasing in the private signal, x. Consequently, the share of the subjects attacking the status quo is decreasing in the strength of the fundamentals, θ. (b) The individual probability of attack and the aggregate size of the attack decrease in the cost of attacking, c.*

This result is consistent with the model's Prediction 1(a). Support for Result 1 comes from Figures 3 and 4 and the regressions in Table 4. Figure 3 plots the individuals' probability of choosing action A for the different ranges of the private signal, x, in the three cost treatments. The figure demonstrates that the probability of attack is decreasing in the private signal over the depicted range of x.

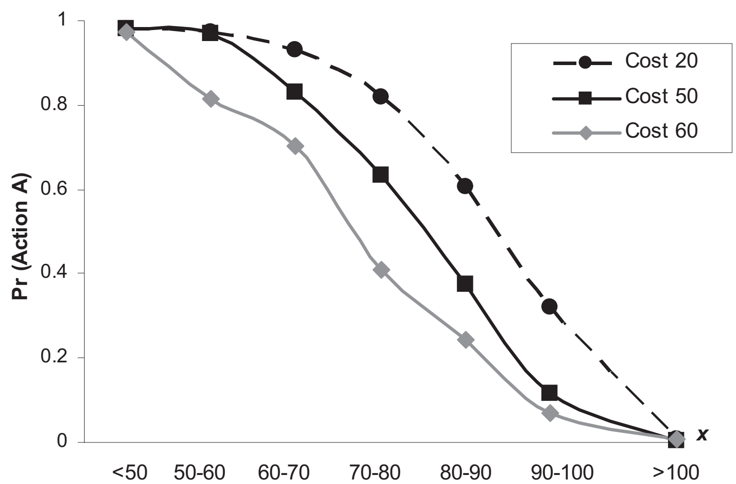


Figure 3: Probability of Attack for Different Realizations of x

Figure 4 reports the size of the attack (i.e., the share of subjects choosing action A) across the different realizations of the parameter θ for all three cost treatments. The figure plots the average size of the attack for the realizations of θ in a given range: in the rounds with low draws of the parameter θ, the size of the attack is close to 1 (everyone choosing action A); in the rounds with high draws of θ, the size of the attack is close to 0 (everyone choosing action B). There is the intermediate range of fundamentals for which the size of the attack is decreasing in θ. The tipping point (the range of θ where the subjects switch between actions A and B) depends on the cost of attacking.

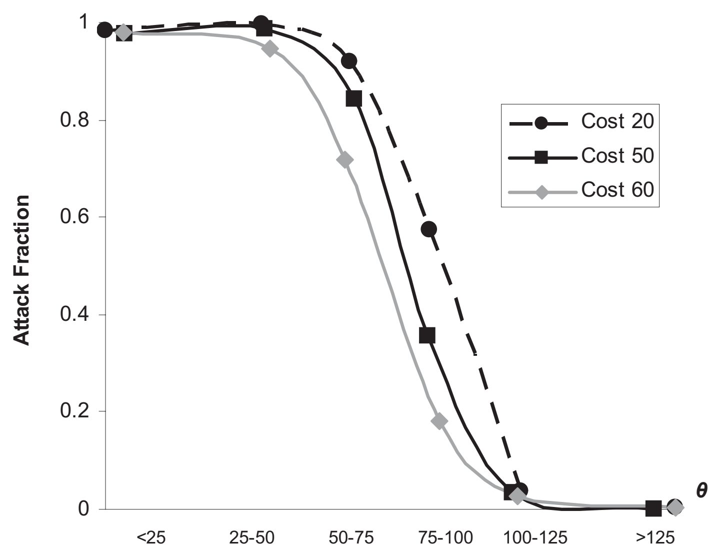


Figure 4: Aggregate Size of Attack for Different Realizations of θ

In addition to showing the negative relationship between the probability of attack and the strength of the fundamental/private signal, Figures 3 and 4 show that the probability of attack is decreasing in the cost of attacking, both at the individual level and on average. This is consistent with Prediction 1(b) of the model. However, the effect of the cost is only marginally statistically significant and small in magnitude, evidenced by the regression analysis in Table 4.

Table 4 reports the results of individual-level regressions of action in stage 1 on the private signal, x, and various controls, including the cost of attacking and the belief about the number of attackers. The dummy variable for the first 10 rounds of the experiment and its interaction terms with x, the cost of attacking, and the belief variables test for any potential order effects that may lead to inter-round learning. We use ordinary least squares (OLS) with subject and round fixed effects to estimate this linear probability model.[14]

[14]The results are qualitatively similar when we use a conditional logit specification (Chamberlain 1980).

Table 4: Determinants of the Individuals' Actions in Stage 1

Variable	Dependent Variable: Action in Stage 1 (OLS)				
	1	2	3	4	5
Private signal, x	-0.007^{***}	-0.007^{***}	-0.001^{***}	-0.001^{***}	-0.001^{**}
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Cost of Attacking	-0.003^{**}	-0.003^{*}	-0.0005^{*}	-0.0006^{*}	-0.0006
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Belief about Size of Attack			0.065^{***}	0.067^{***}	0.068^{***}
			(0.0019)	(0.0024)	(0.0021)
1st 10 Rnds Dummy				0.061	0.065
				(0.069)	(0.067)
x × 1st 10 Rnds				-0.001^{**}	-0.001^{**}
				(0.0003)	(0.0003)
Cost × 1st 10 Rnds				0.0004	0.0004
				(0.0006)	(0.0006)
Belief × 1st 10 Rnds				-0.003	-0.003
				(0.003)	(0.003)
R^2	0.59	0.59	0.83	0.83	0.84
Subject Fixed Effects	No	Yes	No	No	Yes
Round Fixed Effects	Yes	Yes	Yes	No	No
No. of observations	6000	6000	6000	6000	6000

Notes: Robust standard errors in parentheses (clustered at the group level). For sessions 5 and 6, only the no-new-information data are used. Significance levels: * 10%, **5%, ***1%.

The regression analysis confirms that the size of the private signal, x, has a negative, statistically significant effect on the choice of action A. In particular, a 10-point increase in the private signal reduces the individual's probability of attacking by 7 percent (see specifications (1) and (2) in Table 4). This effect remains also significant when controlling for the cost of attacking and for potential effects of learning across rounds. The cost of attacking has a marginally significant negative effect on the probability of attack, which is consistent with Prediction 1(b) of the model, although the coefficients are small in magnitude. Learning over time (rounds) does not have a direct effect on behavior (the coefficient on the dummy for the first ten rounds of each session is not statistically significant).[15] However, we see that the interaction term between the signal and the

[15] The model does not give a prediction about the effect of learning from round to round. Each round

first-ten-rounds dummy is significantly negative. This means that the effect of x on the probability of attack is stronger in the first ten rounds than in the remaining rounds of the session. This implies that subjects are becoming more aggressive over time. Introducing a time trend into the regressions and controlling for the treatment order effects does not significantly change these results (see Table B2 in Appendix B).

Recall that before revealing the outcome of the experiment at a particular stage, we ask each subject about his or her belief about the number of other group members choosing action A. The impact of beliefs on observed behavior is summarized in the following result.

Result 2 **(Stage 1 Beliefs)**. *The individual probability of attacking the status quo (choosing action A) in stage 1 is increasing in the subjects's belief about the number of others who are attacking. Moreover, this belief is decreasing in the private signal, x.*

Support for Result 2 comes from the regression specifications (3) and (4) of Table 4 and the regressions in Table 5. The expectation of others' actions has a significant effect on the subject's own action. In particular, Specifications (3) and (4) of Table 4 show that a subject who expects one more person to participate in the attack in a given round is 7 percent more likely to attack the status quo. This finding supports the strategic complementarity feature of the model and is consistent with Prediction 2. Furthermore, Prediction 2 states that the expected size of the attack decreases in the private signal. The support for this prediction comes from the regression results reported in Table 5. The negative and statistically significantly coefficient on the private signal implies that subjects with lower signals expect others to attack less, as the theory predicts. Once again, the interaction term between the signal and the first-ten-rounds dummy is significantly negative. This means that the effect of x on the expected size of the attack is stronger in the first ten rounds than in the remaining rounds of the session. This implies that subjects are also expecting others to become more aggressive over time.

---

involves a new independent draw of the fundamental θ and therefore represents a new "economy". The model only tells us about learning between stages within each round.

Table 5: Determinants of Individuals' Beliefs about the Number of Attackers in Stage 1

Variable	Dependent Variable: Belief in Stage 1 (OLS)				
	1	2	3	4	5
Private signal, x	-0.085^{***}	-0.084^{***}	-0.079^{***}	-0.078^{***}	-0.015^{***}
	(0.003)	(0.004)	(0.004)	(0.005)	(0.002)
Cost of Attacking	-0.032^{**}	-0.032^{*}	-0.039^{**}	-0.035	0.007
	(0.015)	(0.019)	(0.016)	(0.023)	(0.005)
1st 10 Rnds Dummy			0.114	0.209	-0.062
			(0.760)	(0.850)	(0.276)
x × 1st 10 Rnds			-0.009^{**}	-0.11^{**}	0.002
			(0.004)	(0.005)	(0.001)
Cost × 1st 10 Rnds			0.005	0.005	-0.005
			(0.014)	(0.015)	(0.005)
Actual Attack Size					0.786^{***}
					(0.019)
R^2	0.58	0.63	0.57	0.62	0.81
Subject Fixed Effects	No	Yes	No	Yes	Yes
Round Fixed Effects	Yes	Yes	No	No	No
No. of observations	6000	6000	6000	6000	6000

Notes: Robust standard errors in parentheses (clustered at the group level). For sessions 5 and 6, only the no-new-information data are used. Significance levels: * 10%, **5%, ***1%.

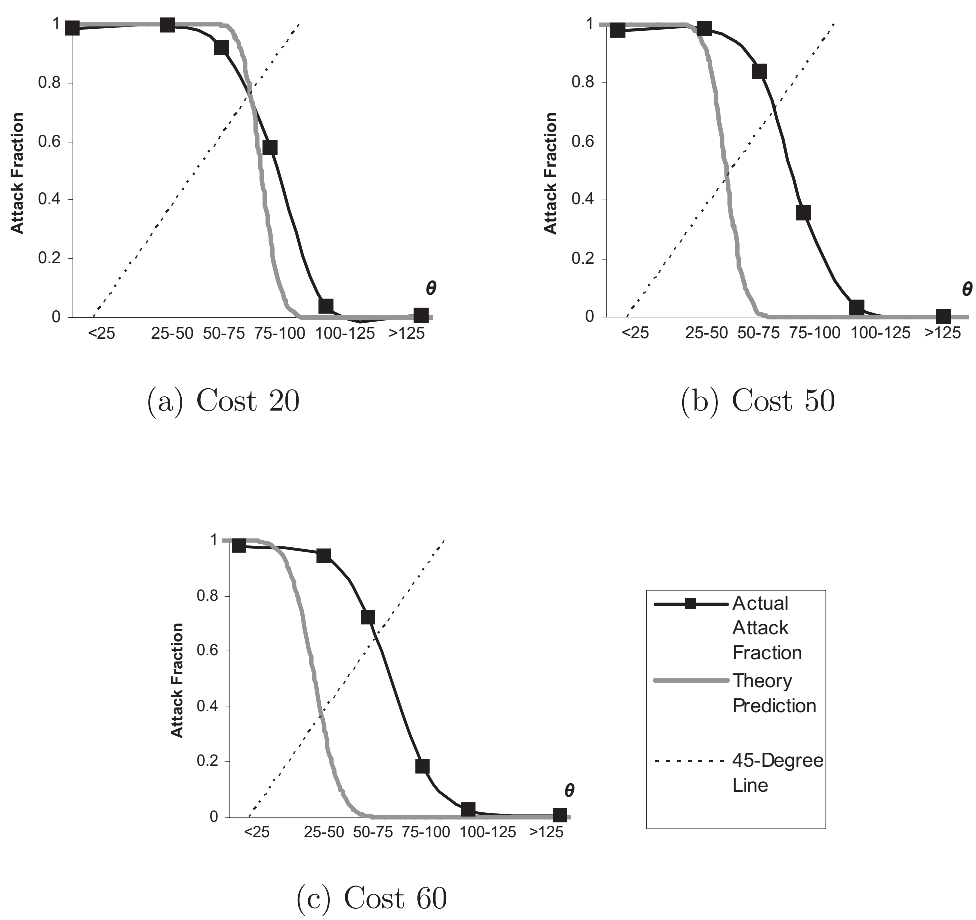
The fact that the private signal affects the subjects' decisions not only directly (Table 4), but also through their expectations about others' strategies (Table 5) serves as evidence that subjects take into account strategic considerations that constitute an important element of the model.

Results 1 and 2 suggest that the data support the qualitative predictions of the model in the first stage. However, these results also raise an important question: does subject behavior deviate from the quantitative predictions? The next result addresses the comparison between actual and predicted actions.

Result 3 **(Stage 1 Excess Aggressiveness)**. *Given the cost of attacking and the realization of the fundamental, θ, the share of subjects attacking the status quo (choosing action A) is larger than the equilibrium share predicted by the theory, particularly in the high-cost treatments. As a consequence, the percentage of successful attacks is significantly* *higher than predicted in the high cost treatments.*

Support for Result 3 comes from Figure 5 which plots the actual size of the attack (same as in Figure 4) and the predicted theoretical size of the attack for different realizations of the fundamental θ in all the cost treatments. The figure also provides a visual representation of the equilibrium for each cost treatment: the theoretical threshold θ^* can be found at the intersection of the aggregate size of the attack and the 45-degree line.

Figure 5 shows that for the high-cost treatments (cost 50 and 60) the fraction of subjects attacking is higher than the theory prediction for the larger range of the fundamental θ.



(a) Cost 20

(b) Cost 50

(c) Cost 60

Figure 5: Comparison of Actual and Predicted Aggregate Size of Attack for Different Realizations of θ

In Figure 5, the average estimated threshold, determined by the intersection between the 45-degree line and the actual size of the attack, is above the theoretical threshold in those cost treatments. In the Cost 20 treatment, subjects' excess aggressiveness is less pronounced and only exists for higher realizations of θ.

An alternative way to view these results is to note that the actual size of the attack is largely insensitive to the variations in the cost of attacking. The evidence for this insensitivity also appears in the regressions in Table 4. While the coefficient on the cost variable of -0.003 is statistically significant at the 10 percent confidence level in most specifications, the magnitude of the effect is small considering that cost varies from 20 to 50 to 60: the probability of attacking decreases only by 3 percent as cost increases by 10 points.

Result 3 raises the question of whether the aggressiveness stems from behavior that is "erroneous" or irrational given the agents' beliefs. To address this question, we postulate the following two hypotheses. The first hypothesis states that subjects are "irrational" in the sense of being intrinsically biased toward attacking: they simply "enjoy" attacking. An alternative hypothesis is that subjects are "rational" in a sense that they follow a best-response strategy given their beliefs of what others will do, and it is their beliefs that are more aggressive than the theory predicts. Result 4 formulates the outcome of the test of these hypotheses.

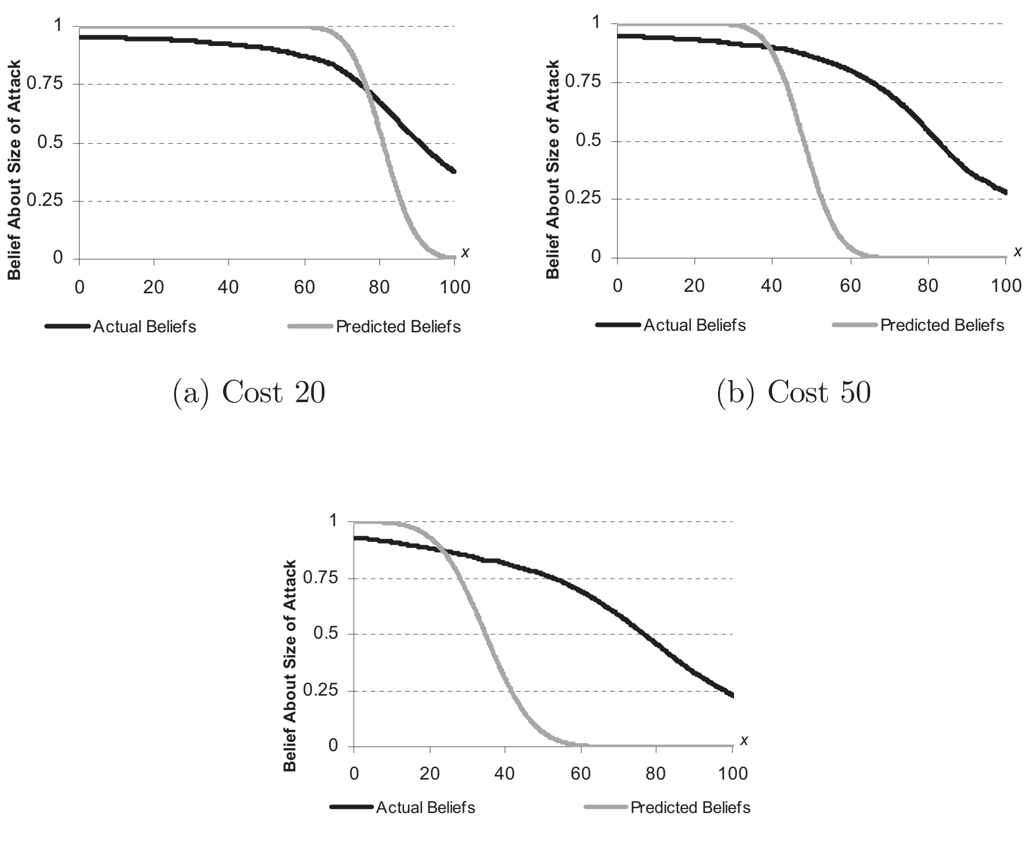
RESULT 4 **(Rationality of Actions)**. *Subjects' above-equilibrium attacking behavior is rational in the sense that subjects are following their best-response strategy given their beliefs about the size of the attack.*

We examine the rationality of subjects' actions for given beliefs by deriving subjects' implicit belief about others' threshold $\widehat{x}$ from their explicit reported belief about the number of attackers (see Appendix A, Section 7.1.4). In other words, after observing each individual's x and belief about the number of attackers, we make the inference on what this particular subject must have believed about others' strategies (estimate a value of $\widehat{x}$). We then compute the expected payoff from attacking given the estimated value of $\widehat{x}$ and the individual's private signal x and compare this payoff to the cost of attacking. A rational agent will attack if and only if the expected payoff from attacking is greater than the cost. We find that subjects are rational in 76.98 percent of cases in the Cost 20 treatment, 90.79 percent of cases in the Cost 50 treatment, and 89.44 percent of cases in the Cost 60 treatment. Thus, the data suggest that, given their aggressive beliefs about others' actions, agents act "rationally" (i.e., in accordance with their individual best-response) in 86 percent of cases, on average.

If most subjects make rational choices for given beliefs, it seems likely that excessively aggressive attacking behavior is driven by excessively aggressive (i.e., optimistic) expectations. Our next result shows that this is indeed the case.

RESULT 5 **(Above Equilibrium Beliefs)**. *For relatively high realizations of x subjects' beliefs about the size of the attack are far more optimistic than the equilibrium belief. This deviation is particularly pronounced in the Cost 50 and the Cost 60 treatments.*

Support for Result 5 comes from Figure 6, in which the actual beliefs about the fraction of agents attacking are estimated from the reported individual belief data using the kernel regression method.



(a) Cost 20

(b) Cost 50

(c) Cost 60

Figure 6: Comparison of Actual and Predicted Beliefs about the Size of Attack for Different Realizations of x

We compare these beliefs to the theoretically predicted beliefs for the realizations of x in the range $[0, 100]$ for Cost 20, Cost 50, and Cost 60 treatments.[16] Figure 6 shows that the subjects' *beliefs* are much more aggressive than predicted by the theory in the

[16]See the Appendix A, Section 7.1.4, for a mathematical formulation of predicted beliefs.

high cost treatments. (The agents' expectations of the size of the attack lie above/to the right of the theoretical expectation for the higher values of x.) These are exactly the cost treatments for which we find more aggressive actions relative to theory (see Figure 5). However, the beliefs in the Cost 20 treatment do not seem to be more aggressive than the theoretical prediction for the realizations of x below 80. Looking back at Figure 5, agents' actions are not more aggressive than the theoretical prediction in the Cost 20 treatment for realizations of x below 80. Note also that Figure 6 demonstrates that the observed beliefs are less sensitive to the individual signal than the theory predicts, since the actual beliefs are everywhere flatter than the predicted beliefs. This feature of the actual beliefs is present in all cost treatments.

Why do the subjects in our experiment exhibit such aggressive expectations about others' behavior? And why are these out-of-equilibrium expectations not corrected over time? One possibility is that subjects' overly aggressive expectations about the others induce them to attack the status quo in an excessively aggressive manner, which then leads to an above equilibrium share of successful attacks. If that were the case, the subjects' above-equilibrium expectations about others' attacking behavior would be reinforced by the experience of an above-equilibrium share of successful attacks. If we compare the actual and the equilibrium shares of successful attacks (see Figure 7), we observe that indeed subjects attack the status quo with considerably higher success rate than predicted. In the Cost 50 treatment, 61 percent of all attacks are successful (compared to 31 percent in theory), while in the Cost 60 treatment, 42 percent of all attacks are successful (compared to 25 percent in theory). Interestingly, the qualitative pattern displayed in Figure 7 also holds for the second 10 rounds of the experiment, indicating that subjects seem to have little reason to revise their aggressive expectations.[17]

[17] In the second 10 rounds, 61 percent of all attacks are successful compared to 32 percent in theory in the Cost 50 treatment and 38 percent of all attacks are successful compared to 21 percent in theory in the Cost 60 treatment. In the Cost 20 treatment, the actual share of successful attacks in the second 10 rounds is 53 percent, which is again not statistically different from the predicted share of 58 percent.

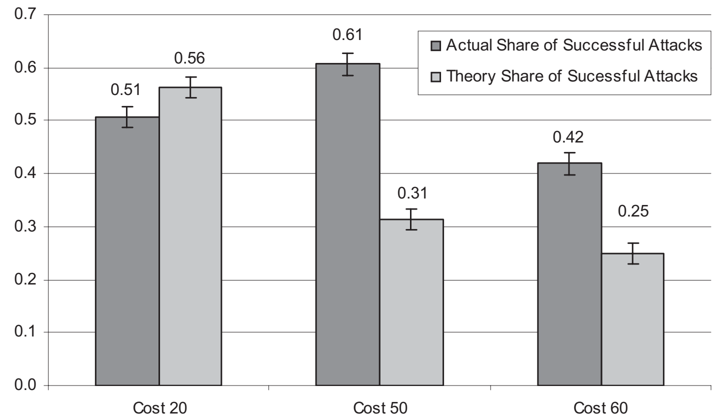
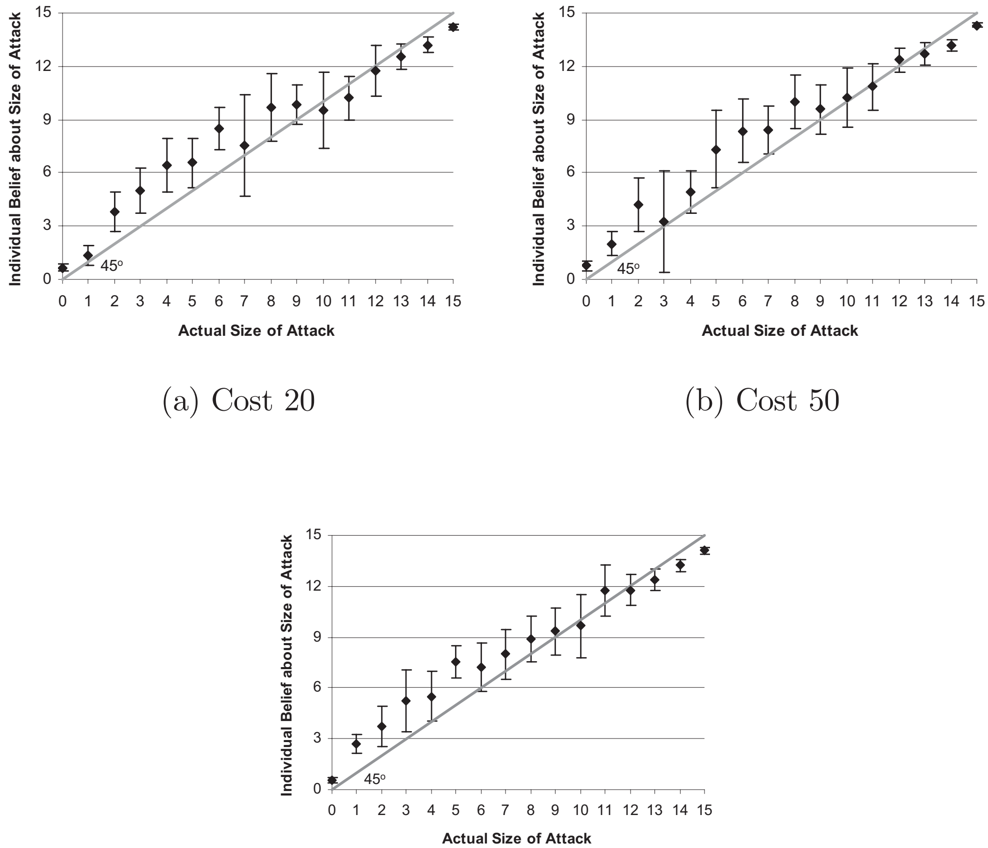


Figure 7: Comparison of Actual and Predicted Shares of Successful Attacks in Stage 1, All Cost Treatments (with 95% Confidence Intervals)

The reasons for the stickiness of subjects' excessively aggressive expectations are examined in more detail in our next result which compares the actual size of the attack with the subjects' expectations about the size of the attack.

RESULT 6 **(Rationality of Beliefs)**. *Subjects' average expectations about the size of the attack are close to the actual size of the attack. Only for small levels of the actual attack size do subjects slightly overestimate others' attacking behavior, whereas for large levels of the actual size of the attack, subjects' expectations are on average rather accurate.*

Support for Result 6 comes from column (5) of Table 5 and Figure 8. The regression specification in column (5) of Table 5 examines the effect of the actual attack size on the individual belief about the size of the attack. The statistically significant coefficient of 0.8 implies that when the actual attack size increases by one attacker, the expected attack size increases, on average, by 0.8 attackers. Moreover, Figure 8 demonstrates that, for the most part, the expected size of the attack deviates very little from the actual size of the attack in all cost treatments.



(a) Cost 20

(b) Cost 50

(c) Cost 60

Figure 8: Individual Beliefs about the Size of the Attack for Different Actual Sizes of the Attack (with 95% Confidence Intervals)

This partially self-fulfilling nature of subjects' beliefs that likely arises due to the presence of strategic complementarity may justify the persistence of excessively aggressive expectations.

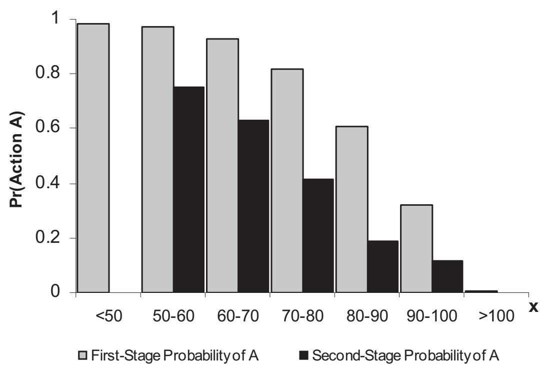
We now ask whether the same conclusions can be drawn once the subjects are allowed to learn about the state of the fundamentals over time.

5.2 Behaviors and Expectations after a Failed Attack

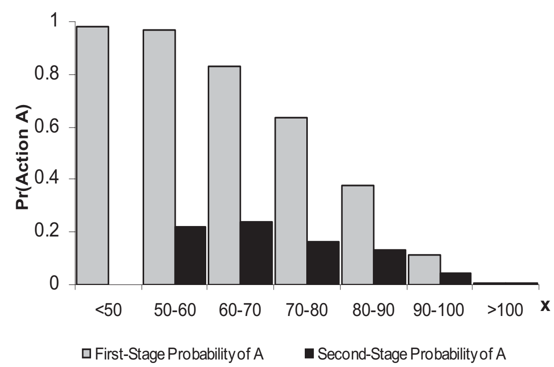
In this section of the paper, we will discuss the effects of learning across stages under the condition that subjects do not receive an additional signal in the second stage.

RESULT 7 **(Stage 2 Behavior, No New Information)**. *If the status quo survives the attack in stage 1, without the arrival of new information, the individual probability* *of attacking (choosing action A) is significantly lower in stage 2 than in stage 1. Consequently, the share of subjects attacking the status quo in stage 2 is greatly reduced relative to the share in stage 1. However, this share of attackers is greater than zero.*

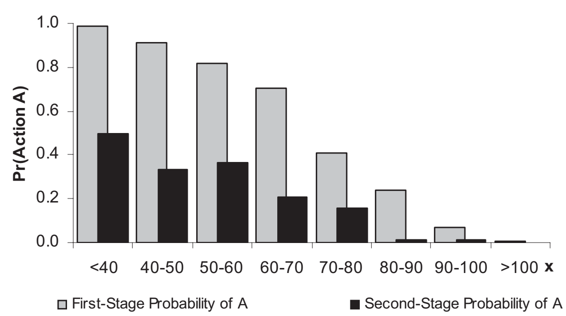
Although second-stage behavior displays excess aggressiveness relative to the stark theoretical prediction of no attacking in the second stage without new information, Result 7 is qualitatively consistent with the theory Prediction 3. Support for Result 7 comes from Figures 9 and 10 and the regressions in Table 6.



(a) Cost 20



(b) Cost 50

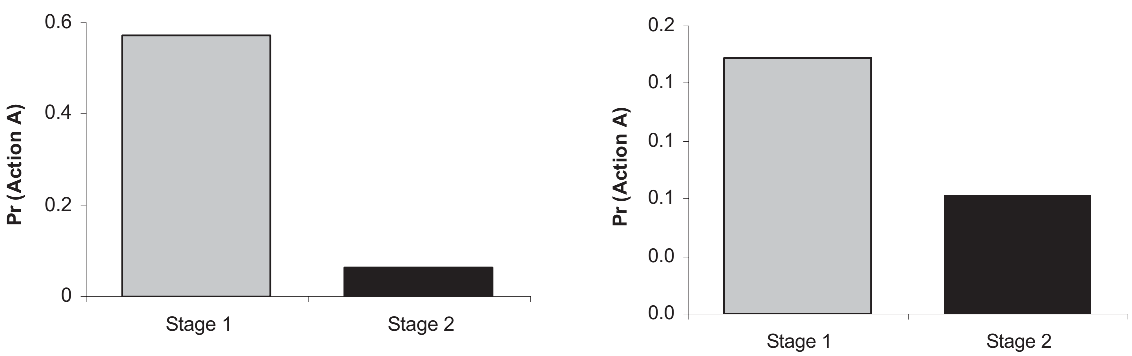


(c) Cost 60

Figure 9: Probability of Attack in Stage 1 and 2 for Different Realizations of x

Figure 9 documents the reduction in individual-level aggressiveness across stages for different realizations of x. The probability of attacking is lower in the second stage relative to the first stage. This evidence is consistent with the theory's prediction regarding the impact of endogenous learning. However, the agents continue to act overly aggressively relative to theory in the second stage of the experiment, because the model predicts that the probability of attacking should be zero without new information. Note also that the reduction in the probability of action A is substantially larger in the high-cost treatments.

Panel (a) of Figure 10 reports the average probability of attack in both stages for all rounds and sessions; Panel (b) uses the data only for those rounds when the game continued into the second stage. In both cases, the average probability of attack in the second stage of the no-new-information treatments is much lower than the average probability of attack in stage one. Moreover, out of all the rounds without new information that continued into stage two, only 3 culminated in a successful attack (all in the Cost 50 treatment). This constitutes a 4.8 percent success frequency in the second stage of the Cost 50 treatment (compared to 61 percent in the first stage), and 0 percent success frequency in the Cost 20 and Cost 60 treatments with no new information (compared to 51 percent and 42 percent in the first stage, respectively; see Figure 7). Thus, we find excess aggressiveness in the second stage (size of the attack is not reduced to zero), but it is not sufficient to trigger successful attacks.



(a) Unconditional Probability of Attack (b) Conditional on Continuing to Stage 2

Figure 10: Average Probability of Action A in Stage 1 and 2
for the No-New-Information Treatments

Table 6 reports the results of a linear probability model estimating the effects of continuation into the second stage on the individual probability of attacking in the case when no new information is revealed to the subjects.[18] Specifications (1) and (2) show that the knowledge that the experiment proceeded into the second stage reduces the probability of attacking by 23 percent. This negative effect is statistically significant

[18]Results do not change significantly if we use the conditional logit specification.

at the 1% confidence level and is qualitatively consistent with Prediction 3 (endogenous learning). However, the 23 percent reduction in aggressiveness is smaller in magnitude than the quantitative prediction of the model, namely that the reduction should be 100 percent.

Table 6: The Impact of a Failed Attack in Stage 1 on Attacking Behavior in Stage 2

Variable	Dependent Variable: Action (OLS)				
	1	2	3	4	5
Private signal, x	-0.006^{***}	-0.006^{***}	-0.001^{***}	-0.001^{***}	-0.001^{**}
	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Cost of Attacking	-0.003^{***}	-0.003^{**}	-0.0005^{*}	-0.0006^{*}	-0.0006
	(0.0008)	(0.0012)	(0.0003)	(0.0004)	(0.0004)
Stage 2 Dummy	-0.242^{***}	-0.227^{***}	-0.040^{***}	-0.049^{***}	-0.040^{***}
	(0.0242)	(0.0258)	(0.0085)	(0.0085)	(0.0087)
Belief about Size of Attack			0.066^{***}	0.064^{***}	0.066^{***}
			(0.0019)	(0.0032)	(0.0023)
1st 10 Rnds Dummy				0.032	0.044
				(0.053)	(0.050)
x × 1st 10 Rnds				-0.0004	-0.0004^{*}
				(0.0002)	(0.0002)
Cost × 1st 10 Rnds				0.0003	0.0002
				(0.0006)	(0.0006)
Belief × 1st 10 Rnds				-0.001	-0.002
				(0.002)	(0.002)
R^2	0.60	0.62	0.85	0.84	0.85
Subject Fixed Effects	No	Yes	Yes	No	Yes
Round Fixed Effects	Yes	Yes	Yes	No	No
No. of observations	8820	8820	8820	8820	8820

Notes: Robust standard errors in parentheses (clustered at the group level). For sessions 5 and 6, only the no-new-information data are used. Significance levels: * 10%, **5%, ***1%.

Specification (3) introduces the subjects' belief about the size of the attack as an explanatory variable for the probability of attacking in both stages. The belief variable is highly statistically significant and positive, which implies that if subjects expect others to attack the status quo, they are more likely to attack as well. Moreover, the inclusion of beliefs reduces the magnitude of the coefficient on the stage variable (although it remains statistically significant and negative). This suggests that endogenous learning affects observed behavior through its effect on the subjects beliefs about others' actions.

Specifications (4) and (5) control for the potential round effects, including interactions between the first ten rounds of the experiment with the explanatory variables. Once again, we see no learning effects between the early and the late rounds of the experiment, with subjects becoming slightly more aggressive over rounds, as the negative coefficient on the interaction term between x and the first ten rounds is marginally statistically significant.

In stage one, we found that subjects' excessively aggressive behavior stemmed from their overly optimistic expectations about others' attacking behavior. Our next result explores whether stage-two excess aggressiveness is also driven by subjects' optimistic beliefs about the size of the attack.

RESULT 8 **(Stage 2 Beliefs, No New Information).** *The knowledge that the first-stage attack was unsuccessful reduces the subjects' expectations about the size of the attack in the second stage without new information. However, the subjects' beliefs remain overly optimistic relative to theory which may once again partly justify their overly aggressive behavior.*

In order to support Result 8, we analyze the impact of a failed attack in the first stage on the subject's individual belief about the size of the attack (see Table B3 in Appendix B). Controlling for the private signal in stage one, the cost of attacking and various round effects, we find that the continuation of the experiment into the second stage reduces the believed number of other attackers by 3 out of 14 (statistically significant at the 1% confidence level). This finding implies that subjects not only reduce own aggressiveness from stage to stage (see Figures 9 and 10), but that they also expect others to exhibit the same kind of learning without arrival of additional private information between stages. The reduction in the aggressiveness of beliefs is consistent with the theory Prediction 4.

However, in all cost treatments, we also find that beliefs are significantly higher than zero (the theory prediction). These aggressive beliefs may partially justify why subjects persist in attacking the status quo in the second stage even with no additional private information. But why do beliefs about the number of attackers remain higher than predicted in the first place? For some of the Cost 50 treatments, the possible reason may be the subjects' ability to coordinate on a few successful attacks in the second stage, which occurred in rounds 10, 11 and 12. However, the overall share of successful attacks out of the total number of rounds that continued into the second stage is small in magnitude, just 4.8 percent (total of 3 successful attacks). In the other cost treatments, we do not observe any successful attacks in the second stage. Aggressive beliefs in the second stage may also result from the feedback that the subjects receive at the end of each round. Recall that at the end of the round, each subject learns whether the attack was successful and the total number of subjects attacking the status quo in each stage. Even if the attack was not successful in the second stage, which should make the subjects more pessimistic about overturning the status quo in subsequent rounds, they also learn that the size of the attack may not have been zero, which could make them relatively more aggressive going forward.

The knowledge that the status quo has survived the stage-one attack greatly reduces the aggressiveness in actions (Figures 9 and 10) and individuals' beliefs in the second stage, albeit not all the way to zero. The findings raise two follow-up questions: (1) Is the endogenous learning effect strong enough to prevent stage-two attacks even as new information arrives? (2) Is there an interaction between subjects' beliefs and learning with new information? We address these issues in the following subsection.

5.3 The Impact of New Information after a Failed Attack

With sufficiently precise new information in the second stage, the theory predicts that an attack becomes possible again. We test this prediction by introducing a treatment condition where subjects receive a new more precise private signal about the fundamental θ in the second stage if the status quo survives the first-stage attack. The first finding is summarized in Result 9.

RESULT 9 **(Stage 2 Behavior, New Information).** *If the status quo survives the attack in stage 1, the individual probability of attacking (choosing action A) is significantly higher with new information than without new information in the second stage. Consequently, the share of the subjects attacking the status quo in stage 2 is greater with new information than without new information.*

Result 9 is qualitatively consistent with Prediction 5 of the model. Support for this result comes from Figures 11 and 12 and Table 7. Figure 11 shows the probability of action A for different realizations of x in the no-new-information (NNI) and the new-information (NI) treatments with cost of action A of 60. For most bins of x, the probability of action A in the second stage for the new-information treatment exceeds the probability of action A for the no-new-information treatment. Moreover, for x above 80, the probability of action A in the second stage of the new-information treatment even exceeds the probability of action A in the first stage. Despite the fact that only one out of 47 attempted attacks

(2.1 percent) was actually successful at overturning the status quo in the second stage of the experiment with new information, subjects persistently attack the status quo more frequently than in the second stage without new information.

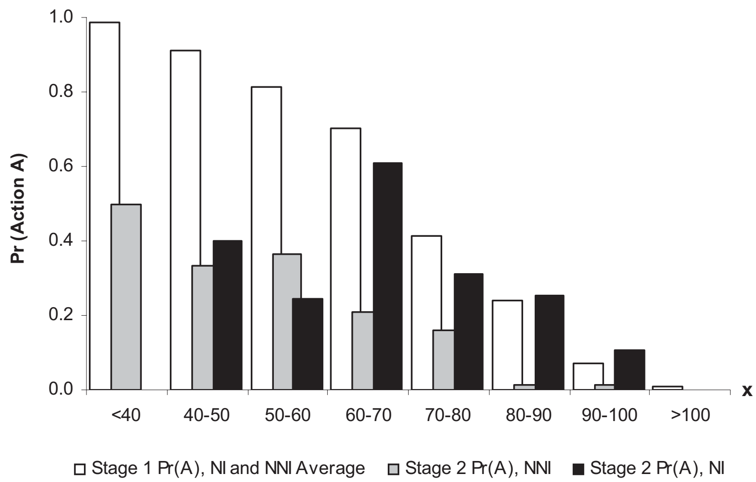


Figure 11: Probability of Action A for Different Realizations of x by Stage for the No-New-Information (NNI) and New-Information (NI) Treatments

The same increase in aggressiveness in the second stage with new information relative to the no-new-information case can be detected by looking at the aggregate level data. Figure 12 contrasts the average probability of action A in the two stages of the experiment for the no-new-information (NNI) and the new-information (NI) treatments. Note that the figure was constructed using only the rounds that continued into the second stage and for which the random number drawn was below 100. This allows us to make the clearest possible comparison between treatments. First, consider the average probability of action A for the two treatment conditions in the first stage. The probabilities are very close in magnitude: 0.25 for the treatment with no new information in stage two and 0.26 for the one with new information in stage two. We can therefore conclude that the anticipation in stage one that new information will arrive in stage two does not alter behavior in stage one. This fact is consistent with the theory. Secondly, note that the probability of action A is reduced dramatically in the no-new-information treatment, as the experiment continues into the second stage, but remains high in the new-information treatment. These results are consistent with the model's predictions in the second stage.

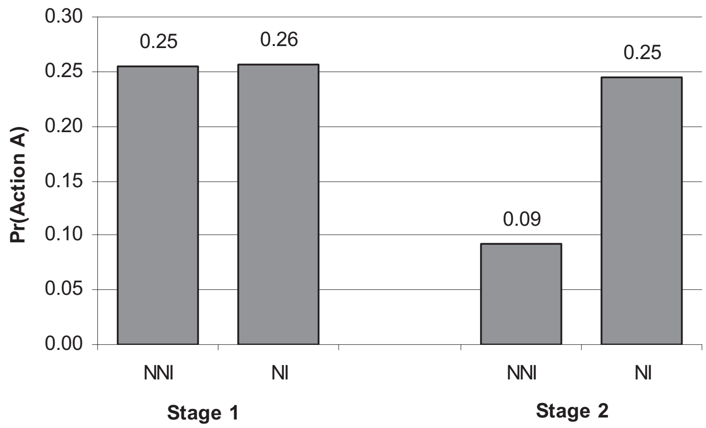


Figure 12: Average Probability of Action A for the Treatments with No New Information (NNI) and New Information (NI)

We confirm the statistical significance of this result by running individual-level OLS regressions of action in the second stage on the private signal, x, and the new-information treatment dummy. Table 7 summarizes the regression results.[19] The coefficient on the new-information (NI) dummy in Specification (1) tells us that subjects are more likely to choose action A in stage two of the new-information treatment than in stage two of the no-new-information treatment, which is consistent with Prediction 5.

[19]The results are robust to using the conditional logit specification. For specifications that include treatment order effects, see Appendix B, Table B4.

Table 7: The Impact of New Information on Stage 2 Behavior

Variable	Dependent Variable: Action in Stage 2 (OLS)		
	1	2	3
Private signal, x	-0.002^{***}	-0.0001	-0.0001
	(0.0005)	(0.0001)	(0.0002)
New Information (NI) Dummy	0.089^{*}	0.017	0.013
(1 if New Information)	(0.048)	(0.020)	(0.013)
Belief about Size of Attack		0.053^{***}	0.050^{**}
		(0.007)	(0.011)
1st 10 Rnds Dummy			-0.020
			(0.073)
x $\times$ 1st 10 Rnds			0.0002
			(0.0003)
NI $\times$ 1st 10 Rnds			-0.009
			(0.017)
Belief $\times$ 1st 10 Rnds			0.006
			(0.009)
R^2	0.21	0.57	0.57
Subject Fixed Effects	Yes	Yes	Yes
Round Fixed Effects	Yes	Yes	No
No. of observations	1395	1395	1395

Notes: Robust standard errors in parentheses (clustered at the group level).
Significance levels: * 10%, **5%, ***1%.

Note that beliefs still have a significant impact on attacking behavior in the second stage with new information. In fact, the inclusion of the belief variable in the regressions (Specifications (2) and (3)) in Table 7 renders the coefficients on the private signal x and the new information dummy insignificant. This implies that both x and the new information dummy impact observed behavior in the second stage of the experiment mainly through their effect on the subjects' beliefs about the actions of others. This fact suggests that subjects' behavior is qualitatively consistent with best responses to their beliefs which brings us to our final result.

Result 10 **(Stage 2 Beliefs, New Information).** *Arrival of new information sig-*

nificantly increases the expected size of the attack in the second stage.

Support for Result 10 comes from running an individual-level regression of the expected size of the attack on the new information dummy in the second stage, controlling for the private signal and round and subjects fixed effects (see Table B5 in Appendix B). The regression indicates that the arrival of a new private signal significantly increases the belief about the number of other subjects attacking. Thus, new information causes the subjects' expectations to become relatively more aggressive in the second stage (Prediction 6), which partly justifies the associated increase in attacking behavior relative to the no-new-information case that we documented in the previous result (see Figures 11 and 12, for example).

In line with Stage 1 evidence, Stage 2 results provide support for the qualitative predictions of the model and further highlight the role played by the perceptions of overall group aggressiveness. Although subjects pay attention to incoming private information and to the costs of attacking, the behavior is consistently overly aggressive relative to equilibrium predictions of the model and can only be rationalized once the subjects' expectations of others' actions are taken into account.

6 Conclusion

This study is motivated by important open questions associated with events that require coordination, such as currency crises, capital flight episodes, or political uprisings. We exploit recent advances in dynamic global game theory to examine the causal impact of the strength of fundamentals, the arrival of new information after a failed attack, and the cost of attacking on subjects' attacking behavior in a laboratory experiment that mimics the essential features of a currency crisis. The clean theory predictions of dynamic global games combined with the opportunity for enhanced control through laboratory experimentation allow us to make a number of conclusions.

As the theory predicts, we find that (i) an increase in the underlying strength of the status quo and (ii) an increase in the cost of attacking cause a reduction in the agents' propensity to attack the status quo. However, reducing the cost does not have as large an effect on attacking behavior as the model predicts. We also demonstrate that, after a failed attack, the knowledge that the status quo has survived greatly decreases the probability that individuals will show attacking behavior. However, if the agents receive new private information after a failed attack in the form of a more precise signal about the strength of the fundamentals, the probability of a new attack increases significantly.

Because we measure subjects' beliefs about other players' attacking behavior, we are able to gain further insights into the behavioral mechanisms behind their actions. In particular, the results suggest that the variations in the strength of the fundamentals, the arrival of new information, and changes in cost affect observed behavior through their impact on subjects' beliefs. For example, the survival of the status quo after a failed attack induces subjects to lower their expectations about the number of other players attacking, which in turn induces them to (rationally) reduce their own probability of attacking.

While these findings provide important qualitative support for the predictions of the theory and the behavioral mechanisms assumed by it, we also find interesting patterns that contradict the theory. In particular, the subjects exhibit considerably more aggressive attacking behavior than the theory predicts. We show that this excess aggressiveness is not simply the result of irrational behavior or a "taste for aggression", but that it stems from subjects' overly aggressive beliefs about the other players' attacking behavior. In fact, given their beliefs, the vast majority of the subjects make attacking decisions that maximize their expected payoffs. Moreover, it is also not possible to argue that subjects simply have the wrong beliefs because, for a large number of actual attack sizes, subjects' beliefs about the attack size do not significantly deviate from the actual levels (Figure 8). Thus, subjects frequently not only believe that others are overly aggressive, but also experience this to be the case. In our view, strategic complementarity is the reason for this partially self-fulfilling nature of subjects' "deviant" beliefs: overly aggressive beliefs induce overly aggressive behavior, which in turn (partially) justifies the initial beliefs. This argument also explains the persistence of the excess aggressiveness of beliefs and behaviors.

In our view, these findings have potentially important implications. One major achievement of the theory of global games is that it often provides clean (i.e., unique) predictions for situations that require coordination, while previous models predicted the existence of multiple equilibria. However, if deviations from the unique equilibrium of a global game are partially self-confirming (because arbitrary beliefs induce attacking behaviors that partially confirm the initial beliefs), the range of behaviors that one can expect to occur may be considerably larger than the range of behaviors the theory predicts. Moreover, the arbitrariness of the initial expectations brings back a problem that resembles the multiple equilibrium problem: it is difficult to make a precise prediction unless one is willing to make an assumption about the initial beliefs. We believe, therefore, that our results may constitute an exciting challenge that may induce the further development of theories of coordination behavior.

7 Appendix A: Model and Derivations (Not for Publication)

This appendix describes the best-reply behavior of the agents and characterizes the equilibrium in both stages of the dynamic game.

7.1 First-Period Predictions

7.1.1 Definitions in the First Period

We begin by introducing several key objects that describe the behavior of the agents in this framework. We first assume that an individual agent i believes that other agents attack the status quo if and only if their private signal is below a certain critical value, denoted by $\widehat{x}_{i1}$ in the first stage of the game.[20] This threshold parametrizes the aggressiveness of the agents in the economy. A higher value of $\widehat{x}$ would mean that agents are willing to attack for a larger range of private signals, signifying a greater degree of overall aggressiveness.

Definition 1 *(1) Let $A(\theta;\widehat{x}_1)$ denote the size of the attack when the state of the fundamentals is θ and the other agents attack if and only if they receive signals $x_1 \leq \widehat{x}_1$.*

(2) Let $\widehat{\theta}(\widehat{x}_1)$ denote the critical value of the fundamentals such that, when other agents attack if and only if they receive signals $x_1 \leq \widehat{x}_1$, the status quo collapses $(R = 1)$ if and only if $\theta \leq \widehat{\theta}_1(\widehat{x}_1)$.

(3) Let $V(x_1;\widehat{x}_1)$ denote the expected payoff of an attacking agent who receives a signal x_1 and thinks that other agents attack if and only if they receive signals below $\widehat{x}_1$.

The size of the attack $A(\theta;\widehat{x}_1)$ is simply the mass of agents who have received signals below $\widehat{x}_1$ which can be written as

$$A(\theta;\widehat{x}_1) = Pr[x_1 \leq \widehat{x}_1|\theta]^{21}$$

Recall that $x_1 = \theta + \sigma_{x1}\varepsilon$ where $\varepsilon \sim N(0,1)$. Thus, $A(\theta;\widehat{x}_1) = Pr[x_1 \leq \widehat{x}_1|\theta] = Pr[\varepsilon \leq \frac{\widehat{x}_1-\theta}{\sigma_{x1}}] = \Phi\left(\frac{\widehat{x}_1-\theta}{\sigma_{x1}}\right)$ which we summarize in the following useful expression:

$$A(\theta;\widehat{x}_1) = \Phi\left(\sqrt{\beta_1}(\widehat{x}_1-\theta)\right)^{22} \tag{1}$$

[20] For now, we suppress the subscripts i for notational tractability.

[21] Note that we suppress the dependence on exogenous parameters z, c, α, and β_1 for notational tractability.

[22] The first equality stems from the fact that $\varepsilon = \frac{\widehat{x}_1-\theta}{\sigma_{x1}}$, the second arises from the properties of the standard normal distribution, and the third can be established when we recall the definition of the

Taking the partial derivatives of equation 1 with respect to θ and $\widehat{x}_1$, we find that the measure of agents attacking is decreasing in θ and increasing in $\widehat{x}_1$. Intuitively, the negative relationship between $A(\theta;\widehat{x}_1)$ and θ means that the size of the attack increases as the fundamentals in the economy get weaker. The positive relationship between $A(\theta;\widehat{x}_1)$ and $\widehat{x}_1$ implies that an increase in the aggressiveness of the agents in the economy (a higher value of $\widehat{x}_1$) leads to a larger share of agents attacking the status quo for all realizations of θ. Note that Figure 2 in section 2 provides a graphical representation of these monotonicities.

The critical value $\widehat{\theta}_1(\widehat{x}_1)$ represents the value of the fundamentals below which the status quo collapses and solves the equation

$$\widehat{\theta}_1(\widehat{x}_1) = \Phi\left(\sqrt{\beta_1}(\widehat{x}_1 - \widehat{\theta}_1(\widehat{x}_1))\right)$$

In Figure 1 in section 2, $\widehat{\theta}_1(\widehat{x}_1)$ can be found at the intersection of the size of the attack, $A(\theta;\widehat{x}_1)$, with the 45-degree line. Thus, a higher level of aggressiveness (higher $\widehat{x}_1$) moves this intersection to the right, which implies a higher value of $\widehat{\theta}_1(\widehat{x}_1)$, so that the regime collapses for a greater range of realizations of θ.

The expected payoff of an attacking agent that receives a signal x_1 and thinks that other agents attack if and only if they receive signals below $\widehat{x}_1$ is given by

$$V(x_1;\widehat{x}_1) = E[U(a_1, A(\theta;\widehat{x}_1), \theta|x_1)]$$

where $U(\cdot)$ represents the utility function defined in Section 2.1. Given that we have defined $\widehat{\theta}_1(\widehat{x}_1)$ such that regime change occurs if and only if $\theta \leq \widehat{\theta}_1(\widehat{x}_1)$, we can use the utility function of the agent to re-write the individual's payoff as

$$V(x_1;\widehat{x}_1) = E[U(a_1, A(\theta;\widehat{x}_1), \theta|x_1] = a_1(yPr[\theta \leq \widehat{\theta}_1(\widehat{x}_1)|x_1] - c)$$

The posterior of the agent about the value of the fundamental θ is given by

$$\theta|x_1, z \sim N\left(\delta_1 x_1 + (1-\delta_1)z, (\alpha+\beta_1)^{-1}\right)$$

precision of private information $\beta_1 = \sigma_{x1}^{-2}$.

where $\delta_1 \equiv \frac{\beta_1}{\beta_1+\alpha}$ is the relative precision of private information. Hence, the posterior probability of regime change is

$$\begin{aligned} Pr[\theta &\leq \widehat{\theta}_1(\widehat{x}_1)|x_1] = 1 - \Phi\sqrt{\alpha+\beta_1}[\delta_1 x_1 + (1-\delta_1)z - \widehat{\theta}_1(\widehat{x}_1)] \\ &= \Phi\left(\sqrt{\alpha+\beta_1}[\widehat{\theta}_1(\widehat{x}_1) - \delta_1 x_1 - (1-\delta_1)z]\right) \end{aligned}$$

Substituting this expression for $Pr[\theta \leq \widehat{\theta}_1|x_1]$ into $V(x_1, \widehat{x}_1)$ and setting $a_1 = 1$ obtains another useful expression:

$$V(x_1; \widehat{x}_1) = y\Phi\left(\sqrt{\alpha+\beta_1}[\widehat{\theta}_1(\widehat{x}_1) - \delta_1 x_1 - (1-\delta_1)z]\right) - c. \tag{2}$$

Taking the partial derivatives of equation (2) with respect to x_1 and $\widehat{x}_1$, we find that the expected payoff from attacking the status quo is decreasing in the private signal x_1 and increasing in $\widehat{x}_1$. Intuitively, the agent's payoff from attacking is decreasing in the private signal since a higher value of x_1 implies that there is less of a chance that the fundamentals are weak. On the other hand, the payoff from attacking is increasing in $\widehat{\theta}_1$, because a higher critical value below which the regime collapses means that the attack is successful more often.

We now turn to best-reply behavior of the agents given their information about the state of the economic fundamentals.

Definition 2 *Let $BR(\widehat{x}_1)$ denote the value of x_1 that sets $V(BR; \widehat{x}_1) = 0$.*

Lemma 1 *The best-response strategy of an agent is to attack if and only if $x_1 \leq BR(\widehat{x}_1)$.*
Proof. *$x_1 \leq BR(\widehat{x}_1) \Longrightarrow V(x_1, \widehat{x}_1) > 0 \Longrightarrow$ it is dominant to attack.* ■

Intuitively, receiving the signal $BR(\widehat{x}_1)$ makes the marginal agent indifferent between attacking and not attacking the status quo. For all signals below $BR(\widehat{x}_1)$, the gross payoff from attacking outweighs the cost of attacking. Thus, the agent's best response is to attack the status quo. For all signals above, the gross payoff from attacking is below the cost, which implies that the agent's best-response is to refrain from attacking.

The relationship between $BR(\widehat{x}_1)$, $V(\cdot)$, and x_1 is illustrated in Figure A1 below. Note also that $BR(\widehat{x}_1)$ is a function of other agents' strategy $\widehat{x}_1$. An increase in $\widehat{x}_1$ (i.e., a rise in the other agents' aggressiveness level) would raise the agent's own payoff from attacking. This, in turn, will lead to an increase in $BR(\widehat{x}_1)$: the agent will become herself more aggressive, if she believes that others are being more aggressive. This is the nature of strategic complementarity that is at the heart of this model and that we will hereafter refer to as "best-response reasoning."

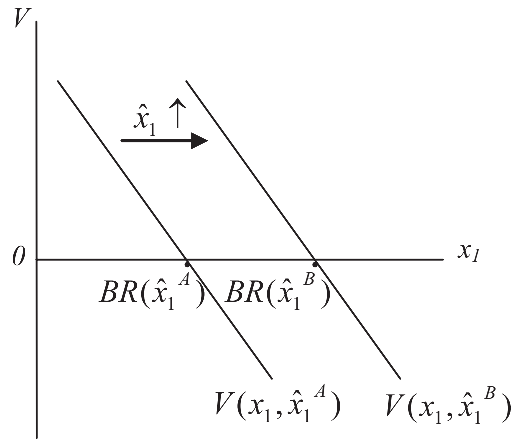


Figure A1: The Payoff Function $V(\cdot)$
for Different Realizations of the x_1 and
Comparative Statics on $\widehat{x}_1$

7.1.2 Monotone Equilibria in the First Period

Let us first focus on the equilibrium outcomes in the first period of the game. Note that it is strictly dominant to attack for sufficiently low signals, namely for $x_1 < \underline{x}$, where $\underline{x}$ solves $Pr(\theta \leq 0|\underline{x}) = c/y$. Similarly, it is strictly dominant not to attack for sufficiently high signals, namely for $x_1 > \overline{x}$, where $\overline{x}$ solves $Pr(\theta \leq 1|\overline{x}) = c/y$. This suggests that we should look for monotone equilibria, that is, equilibria in which $a(x)$ is monotonic in x.

In a monotone equilibrium, for any realization of z, there exists a threshold $x_1^*(z)$ such that an agent attacks if and only if $x \leq x_1^*(z)$. Furthermore, as we know, the measure of agents attacking is decreasing in θ, and therefore there exists a threshold $\theta_1^*(z)$ such that the status quo is abandoned if and only if $\theta \leq \theta_1^*(z)$. A monotone equilibrium corresponds to the pair (x_1^*, θ_1^*).[23]

Using the above definitions, we first characterize the equilibrium threshold θ_1^* for a given x_1^*. For given realizations of θ and z, the aggregate size of the attack is given by the mass of agents with signals $x_1 \leq x_1^*$. Since we have shown that $A(\theta)$ is decreasing in θ, regime change occurs if and only if $\theta \leq \theta_1^*$, where θ_1^* solves $\theta_1^* = A(\theta_1^*)$, or equivalently

$$\theta_1^* = \Phi(\sqrt{\beta_1}(x_1^* - \theta_1^*)). \tag{3}$$

Graphically, the equilibrium threshold θ_1^* is represented by the intersection of $A(\theta)$ with

[23] Again, we drop the dependence on z for notational tractability.

the 45-degree line.

We next characterize the equilibrium threshold x_1^* for a given θ_1^*. Given that regime change occurs if and only if $\theta \leq \theta_1^*$, the payoff of the agent is $V(\cdot)$ which we have shown to be decreasing in x_1. It follows that the agent attacks if and only if $x_1 \leq x_1^*$, where x_1^* solves

$$x_1^* = BR(x_1^*)$$

Explicitly, this condition can be re-written as

$$\Phi\left(\sqrt{\beta_1+\alpha}(\theta_1^* - \frac{\beta_1}{\beta_1+\alpha}x_1^* - \frac{\alpha}{\beta_1+\alpha}z)\right) = c/y. \tag{4}$$

A solution to the system of two equations (3) and (4), (x_1^*, θ_1^*), always exists and is unique for all z if and only if $\beta_1 \geq \frac{\alpha^2}{2\pi}$, which we show below.[24]

A direct implication of the equilibrium in the first period is that a decrease in the cost of attacking, c, increases each agent's incentive to attack. Moreover, the lower cost also causes every agent to expect others to attack more. In other words, a decrease in the cost of attacking is isomorphic to an increase in the aggressiveness of the agents (an increase in $\widehat{x}_1$), in which case the above comparative statics apply (see Figure 2 in section 2).[25]

7.1.3 Properties of the Equilibrium in the First Period

In this section, we establish several relationships and derive certain properties of the equilibrium objects that prove to be useful for the interpretation of the model. First, we solve for the equilibrium threshold in period 1, $\theta_1^*(z)$, for the special case where cost of attacking, c, equals $\frac{1}{2}$ (as was the case in some of the experimental treatments)

$$\theta_1^* = \Phi(\sqrt{\beta_1}(x_1^* - \theta_1^*)). \tag{5}$$

Solving (5) for x_1^*, we get

$$x_1^* = \theta_1^* + \beta_1^{-1/2}\Phi^{-1}(\theta_1^*).$$

[24] Note that iterated elimination of strictly dominated strategies ensures that, when the monotone equilibrium is unique, there is no other equilibrium.

[25] Note that in a coordination game with complete information, such as Obstfeld (1996), the cost of attacking plays no role in equilibrium play: either everyone attacks the status quo or no one attacks the status quo, regardless of cost. Moreover, agents' strategies need not be monotonic in θ.

Now we substitute x_1^* into the other equilibrium equation, namely

$$\Phi\left(\sqrt{\beta_1+\alpha}(\theta_1^* - \frac{\beta_1}{\beta_1+\alpha}x_1^* - \frac{\alpha}{\beta_1+\alpha}z)\right) = 1/2$$

$$\Phi\left(\sqrt{\beta_1+\alpha}(\theta_1^* - \frac{\beta_1}{\beta_1+\alpha}\left[\theta_1^* + \beta_1^{-1/2}\Phi^{-1}(\theta_1^*)\right] - \frac{\alpha}{\beta_1+\alpha}z)\right) = 1/2.$$

Putting the terms that contain θ_1^* on the left-hand side of the expression, we get

$$\frac{\alpha}{\beta_1+\alpha}\theta_1^* - \frac{\beta_1^{1/2}\Phi^{-1}(\theta_1^*)}{\beta_1+\alpha} = \frac{\Phi^{-1}(1/2)}{\sqrt{\beta_1+\alpha}} + \frac{\alpha}{\beta_1+\alpha}z. \tag{6}$$

We can simplify this expression further, using the fact that $\Phi^{-1}(1/2) = 0$ and multiplying both sides by $\frac{\beta_1+\alpha}{\alpha}$

$$\theta_1^* - \frac{\beta_1^{1/2}}{\alpha}\Phi^{-1}(\theta_1^*) = z. \tag{7}$$

Solving for θ_1^* gives us the equilibrium threshold in the first round, $\theta_1^*(z)$, such that the regime will collapse in the first round ($R_1 = 1$) if and only if $\theta \leq \theta_1^*$. Such a solution exists and is unique if and only if the following relationship holds for the precisions: $\beta_1 \geq \alpha^2/(2\pi)$. To see this, we define

$$G(\theta_1^*(z), z) \equiv z - \theta_1^* + \frac{\beta_1^{1/2}}{\alpha}\Phi^{-1}(\theta_1^*) = 0.$$

Note that $G(\theta_1^*(z), \cdot)$ is continuous and differentiable in $\theta \in (0,1)$, and that $G(0,z) = -\infty$ and $G(1,z) = \infty$, which implies that there necessarily exists a solution and any solution satisfies $\theta_1^*(z) \in (0,1)$. This establishes existence. To prove uniqueness, note that

$$\frac{\partial G(\theta_1^*(z), z)}{\partial \theta_1^*} = \frac{\beta_1^{1/2}}{\alpha}(\Phi^{-1})'(\theta_1^*) - 1.$$

We can re-write this using the formula for the derivative of an inverse function:

$$\left(\Phi^{-1}\right)'(\theta_1^*) = \frac{1}{\Phi'(\Phi^{-1}(\theta_1^*))} = \frac{1}{\phi(\Phi^{-1}(\theta_1^*))},$$

where $\phi(\cdot)$ is the p.d.f. of the standard normal distribution and is bounded by $\frac{1}{\sqrt{2\pi}}$ (i.e. $\max_{\omega\in\mathbb{R}} \phi(\omega) = \frac{1}{\sqrt{2\pi}}$). Therefore,

$$\frac{\partial G(\theta_1^*(z), z)}{\partial \theta_1^*} = \frac{\beta_1^{1/2}}{\alpha}\frac{1}{\phi(\Phi^{-1}(\theta_1^*))} - 1 > \frac{\beta_1^{1/2}}{\alpha}\sqrt{2\pi} - 1.$$

Then if $\frac{\beta_1^{1/2}}{\alpha}\sqrt{2\pi} - 1 \geq 0$, or if $\frac{\beta_1}{\alpha^2} \geq \frac{1}{2\pi}$, the function G is strictly increasing in θ_1^* ($\frac{\partial G(\theta_1^*(z),z)}{\partial \theta_1^*} > 0$), which implies a unique solution to (7).

We can use the Implicit Function Theorem to demonstrate that the threshold $\theta_1^*(z)$ is monotonically decreasing in z. Let $F(\theta_1^*(z), z)$ be defined as

$$F(\theta_1^*(z), z) \equiv \frac{\alpha}{\beta_1 + \alpha}\theta_1^* - \frac{\beta_1^{1/2}\Phi^{-1}(\theta_1^*)}{\beta_1 + \alpha} - \frac{\alpha}{\beta_1 + \alpha}z = 0$$

$$\frac{\partial \theta_1^*}{\partial z} = -\frac{\partial F/\partial z}{\partial F/\partial \theta_1^*} = -\frac{-\frac{\alpha}{\beta_1+\alpha}}{\frac{\alpha}{\beta_1+\alpha} - \frac{\beta_1^{1/2}}{\beta_1+\alpha}(\Phi^{-1})'(\theta_1^*)} = \frac{1}{1 - \frac{\beta_1^{1/2}}{\alpha}(\Phi^{-1})'(\theta_1^*)}. \tag{8}$$

As we have shown above, the derivative of the inverse of the c.d.f. of the standard normal is positive and reaches its minimum at $\sqrt{2\pi}$. We also know that the relationship between the precisions is $\frac{\beta_1^{1/2}}{\alpha} \geq \frac{1}{\sqrt{2\pi}}$. Thus, the whole fraction in (8) is negative (i.e., $\frac{\partial \theta_1^*}{\partial z} \leq 0$). Intuitively, θ_1^* is decreasing in z because when the public signal (z) has a high mean, the fundamentals are relatively good. So, the region where the attack will be successful in the first period is relatively small. Thus, the threshold theta is low. In other words, when the mean of the prior is high, the agents are initially pessimistic about their ability to overthrow the regime. So, in the first period, the size of the attack is relatively small. Then, in the second period, if the agents get a sufficiently precise private signal, an attack becomes possible. (That is, agents can become optimistic about their ability to change the status quo.) This is why this scenario can lead to multiplicity.

We can also verify that $\theta_1^*(1/2) = 1/2$ (but only if the public signal is completely uninformative relative to the private signal, that is if $\alpha/\beta_1 \to 0$). Let us substitute 1/2 into equation (6):

$$\frac{\alpha}{\beta_1 + \alpha}\left(\frac{1}{2}\right) - \frac{\beta_1^{1/2}\Phi^{-1}(1/2)}{\beta_1 + \alpha} = \frac{\Phi^{-1}(1/2)}{\sqrt{\beta_1 + \alpha}} + \frac{\alpha}{\beta_1 + \alpha}\left(\frac{1}{2}\right)$$

$$-\frac{\beta_1^{1/2}}{\beta_1 + \alpha} = \frac{1}{\sqrt{\beta_1 + \alpha}}.$$

Squaring both sides

$$\frac{\beta_1}{(\beta_1 + \beta)^2} = \frac{1}{\beta_1 + \beta}$$

$$\frac{1}{1 + \beta/\beta_1} = 1.$$

which is true for all $\beta_1$if $\alpha/\beta_1 \to 0$ or equivalently if $\beta_1/\alpha \to \infty$. To put this differently, we just found the Morris-Shin limit threshold. The Morris-Shin limit is the limit as the ratio of precisions of private and public information approaches infinity, or in other words private information becomes infinitely precise relative to public information. It is

$$\lim_{\frac{\beta_1}{\alpha}\to\infty} \theta_1^*(z) = \frac{1}{2} = 1 - c \equiv \theta_\infty.$$

Finally, we employ the Implicit Function Theorem again to show that $\theta_1^*(z)$ is monotonic in β_1. Define

$$H(\theta_1^*(z), z) \equiv \theta_1^* - \frac{\beta_1^{1/2}}{\alpha}\Phi^{-1}(\theta_1^*) - z = 0$$

$$\frac{\partial\theta_1^*}{\partial\beta_1} = -\frac{\partial H/\partial\beta_1}{\partial H/\partial\theta_1^*} = \frac{\frac{1}{2\alpha}\beta_1^{-1/2}\Phi^{-1}(\theta_1^*)}{1 - \frac{\beta_1^{1/2}}{\alpha}(\Phi^{-1})'(\theta_1^*)}. \tag{9}$$

We already know that the denominator of (9) is always negative. Let us focus on the numerator. The inverse c.d.f. of the normal has the following property:

$$\begin{aligned} \Phi^{-1}(\theta_1^*) &< 0 \text{ if } \theta_1^* < 1/2 \\ \Phi^{-1}(\theta_1^*) &> 0 \text{ if } \theta_1^* > 1/2. \end{aligned}$$

Therefore, to sign this fraction, we need to consider our two cases. In Case $2'$, z is low ($z < 1/2$), which implies that $\theta_1^*(z) > 1/2$ because we have proved above that $\theta_1^*(z)$ is monotonically decreasing in z and that $\theta_1^*(1/2) = 1/2$. This implies that $\Phi^{-1}(\theta_1^*) > 0$, and the numerator of (9) is positive (i.e., $\theta_1^*(z)$ is decreasing in β_1). In Case $2''$, z is high, that is $z > 1/2$, so we know that $\theta_1^*(z) < 1/2$. In this case, $\Phi^{-1}(\theta_1^*) < 0$, the numerator of (9) is negative (i.e., $\theta_1^*(z)$ is increasing in β_1). This shows that $\theta_1^*(z)$ is monotonic in β_1.

7.1.4 Best-Response Behavior and Beliefs in the First Period

Next, we explore best-response behavior of agents given their beliefs. If the data allowed us to observe directly the agent's implicit belief about the probability that the size of the attack is large enough, $Pr[A(\theta; \widehat{x}_1) \geq \theta | x_1]$, we could easily test the prediction that this belief is decreasing in the private signal x_1 and increasing in the other agents' aggressiveness, $\widehat{x}_1$. However, asking for an estimate of the probability of the attack being successful given the private signal seems to be experimentally unrealistic. Thus, in practice, each agent reports his or her belief about the size of the attack.

Definition 3 *Let $m(x_1; \widehat{x}_1)$ denote the belief about the measure of other agents attacking for an agent with signal x_1 who thinks that others attack if and only if $x_1 \leq \widehat{x}_1$.*

This belief is given by the expected size of the attack:

$$m(x_1; \widehat{x}_1) = E[A(\theta; \widehat{x}_1) \geq \theta | x_1] \tag{10}$$

Note that, since $A(\theta; \widehat{x}_1) = \Phi\left(\sqrt{\beta_1}(\widehat{x}_1 - \theta)\right)$, we know that $m(x_1; \widehat{x}_1)$ is decreasing in x_1 and increasing in $\widehat{x}_1$. This result is summarized in Prediction 2, section 2.2.

Using this definition, we construct a criterion for testing the rationality of subjects given their beliefs. For each subject i, we can observe the following: her private signal x_{i1}, her action a_{i1}, and her belief m_{i1} (defined above). However, we do not observe each subjects' notion of the aggressiveness of others, $\widehat{x}_{i1}$. Nonetheless, given x_{i1} and m_{i1}, we can derive an estimate of $\widehat{x}_1$ using the data on the empirically observed beliefs and equations (10) and (1). In particular, equation (10) gives beliefs as a function of the size of the attack, while equation (1) gives the aggregate size of the attack given subjects' beliefs about $\widehat{x}_1$. Given each individual's x and her observed belief m, we make the inference on what the subject must have believed about others' strategies ($\widehat{x}$). We then compute the expected payoff from attacking given the estimated value of $\widehat{x}_1$ and the individual's private signal x_1 and compare this payoff to the cost of attacking. A rational agent will attack (observe the choice $a_{i1} = 1$) if and only if the expected payoff from choosing it is greater than the cost (choose A if and only if $V(x_1; \widehat{x}_1) > 0$).

7.2 Second-Period Predictions

7.2.1 Definitions

Recall that the game continues into the second period if and only if the status quo is in place. If this is the case, we can define the relevant objects in the second period. Though agent i does not observe the measure of agents attacking in the first period, she retains her belief that other agents attacked the status quo in the first period if and only if their private signals were below $\widehat{x}_{i1}$. In addition, she forms a new belief about the aggressiveness of others in the second period which need not be the same as the first-period perceived aggressiveness level. In particular, the agent believes that others will attack the status quo in the second period if and only if their private signals are below $\widehat{x}_{i2}$.[26]

[26] From now on, we again suppress the subscripts i for notational tractability.

Definition 4 *(1) Let $A(\theta;\widehat{x}_2)$ denote the size of the attack when the state of the fundamentals is θ and the other agents attack if and only if they receive signals $x_2 \leq \widehat{x}_2$ in the second period.*

(2) Let $\widehat{\theta}_2(\widehat{x}_2)$ denote the critical value of the fundamentals such that, when other agents attack if and only if they receive signals $x_2 \leq \widehat{x}_2$, the status quo collapses in the second period $(R = 1)$ if and only if $\theta \leq \widehat{\theta}_2(\widehat{x}_2)$.

(3) Let $V(x_2;\widehat{x}_2,\widehat{x}_1)$ denote the second-period expected payoff of an attacking agent who receives a signal x_2 , thinks that other agents attack in the second period if and only if they receive signals below $\widehat{x}_2$, and who thought that other agents attacked in the first period if and only if they received signals below $\widehat{x}_1$.

The size of the attack $A(\theta;\widehat{x}_2)$ is simply the mass of agents who have received signals below $\widehat{x}_2$ which, analogously to the first period, can be written as $A(\theta;\widehat{x}_2) = Pr[x_2 \leq \widehat{x}_2|\theta]$

$$A(\theta;\widehat{x}_2) = \Phi\left(\sqrt{\beta}(\widehat{x}_2 - \theta)\right) \tag{11}$$

The measure of agents attacking the status quo in the second period is decreasing in θ and increasing in $\widehat{x}_2$.

Similarly to the first period, $\widehat{\theta}_2(\widehat{x}_2)$ solves the equation

$$\widehat{\theta}_2(\widehat{x}_2) = \Phi\left(\sqrt{\beta}(\widehat{x}_2 - \widehat{\theta}_2(\widehat{x}_2))\right)$$

Finally, the expected payoff from attacking is given by the posterior probability of a successful attack in the second period, which can be written as

$$V(x_2;\widehat{x}_2,\widehat{x}_1) = y\frac{\Phi\left(\sqrt{\alpha+\beta_2}[\widehat{\theta}_2(\widehat{x}_2) - \delta x_2 - (1-\delta)z]\right)}{\Phi\left(\sqrt{\alpha+\beta_2}[\widehat{\theta}_1(\widehat{x}_1) - \delta x_2 - (1-\delta)z]\right)} - c. \tag{12}$$

The important difference between the first-period payoff $V(x_1;\widehat{x}_1)$ and the second-period payoff $V(x_2;\widehat{x}_2,\widehat{x}_1)$ is represented by the denominator of the above fraction. It arises from the fact that the posterior about θ is truncated in the second period for all the realizations of θ below $\widehat{\theta}_1(\widehat{x}_1)$.[27] Taking respective partial derivatives of $V(x_2;\widehat{x}_2,\widehat{x}_1)$ shows that the expected payoff from attacking is decreasing in the private signal x_2, increasing in the second-period aggressiveness of other agents, $\widehat{x}_2$, and decreasing in the first-period aggressiveness of other agents, $\widehat{x}_1$. The intuition for the last monotonicity is as follows: suppose that the agent suddenly starts to think that other agents attacked in the

[27] Recall that $\widehat{\theta}_1(\widehat{x}_1)$ was the first-period threshold.

first period for a greater range of private signals, i.e., $\widehat{x}_1$ increases. The agent therefore revises her belief about the strength of the fundamentals upwards, since the regime must have been strong enough to have survived a larger first-period attack. Holding x_2 and $\widehat{x}_2$ constant, this makes the agent strictly less likely to want to attack the status quo in the second period.

We now turn to the best-reply behavior of the agents in the second period given their information about the state of the economic fundamentals.

Definition 5 *Let $BR(\widehat{x}_2, \widehat{x}_1)$ denote the value of x_2 that sets $V(BR; \widehat{x}_2, \widehat{x}_1) = 0$.*

Similarly to the first period, we note that the best-response strategy of an agent is to attack if and only if $x_2 \leq BR(\widehat{x}_2, \widehat{x}_1)$ since this would imply that the payoff from attacking in the second period is positive. Intuitively, receiving the signal $BR(\widehat{x}_2, \widehat{x}_1)$ makes the marginal agent indifferent between attacking and not attacking the status quo. For all signals below $BR(\widehat{x}_2, \widehat{x}_1)$, the gross payoff from attacking outweighs the cost of attacking. Thus, the agent's best response is to attack the status quo. For all signals above, the gross payoff from attacking is below the cost, which implies that the agent's best-response is to refrain from attacking.

Note that an increase in the contemporaneous level of aggressiveness, $\widehat{x}_2$, increases the agent's payoff from attacking and therefore leads to an increase in $BR(\widehat{x}_2, \widehat{x}_1)$. That is, the agent becomes relatively more aggressive, believing that others have become more aggressive in the second period. This effect parallels the relationship between $V(x_1; \widehat{x}_1)$ and $\widehat{x}_1$ in the first period. On the other hand, if the agent believes that others were relatively more aggressive in the first period (an increase in $\widehat{x}_1$), the payoff from attacking in the second period decreases which leads to a decrease in $BR(\widehat{x}_2, \widehat{x}_1)$. That is, if the agent believes that other agents were relatively more aggressive in the first period, the observation that the regime has still survived makes the agent relatively less aggressive in the second period.

7.2.2 Monotone Equilibria in the Second Period

We will consider two possibilities for the information structure in the second period. First, suppose that the agents receive no additional information outside the game. In this case, when agents arrive at the second period, they observe that the status quo has survived the first-period attack. From this observation, the agents can infer that the state of the fundamentals is not too weak, because otherwise the status quo would have collapsed under the first attack. In particular, it is now common knowledge that θ is above θ_1^*.

This knowledge causes a first-order-stochastic-dominance shift of beliefs upwards, causing agents' behavior to become less aggressive. This effect, in turn, guarantees that no agent is willing to attack in the second period. We refer to the effect on behavior of the continuation of the regime into the second period without new information as endogenous learning.

Next, we change the information structure in the second period, such that agents receive an additional signal that is sufficiently precise. In particular, the agents receive a signal, $x_{i2} = \theta + \xi_{i2}$, where $\xi_{i2} \sim N(0, 1/\beta_2)$ and β_2 is sufficiently high. We refer to the effects of the arrival of new information in the second period as exogenous learning.

Using the second-period definitions, we first characterize the equilibrium threshold θ_2^* for a given x_2^*. For given realizations of θ and z, the aggregate size of the attack is given by the mass of agents with signals $x_2 \leq x_2^*$. Since we have shown that $A(\theta)$ is decreasing in θ, regime change occurs in the second period if and only if $\theta \leq \theta_2^*$, where θ_2^* solves θ_2^* $= A(\theta_2^*)$, or equivalently

$$\theta_2^* = \Phi(\sqrt{\beta_2}(x_2^* - \theta_2^*)) \tag{13}$$

We next characterize the equilibrium threshold x_2^* for a given θ_2^*. Given that regime change occurs if and only if $\theta \leq \theta_2^*$, the payoff of the agent is $V(\cdot)$ which is decreasing in x_2. It follows that the agent attacks if and only if $x_2 \leq x_2^*$, where x_2^* solves

$$x_2^* = BR(x_2^*, x_1^*)$$

Explicitly, this condition can be re-written as

$$\frac{\Phi(\sqrt{\beta_2 + \alpha}(\theta_2^* - \frac{\beta_2}{\beta_2+\alpha}x_2^* - \frac{\alpha}{\beta_2+\alpha}z))}{\Phi(\sqrt{\beta_2 + \alpha}(\theta_1^* - \frac{\beta_2}{\beta_2+\alpha}x_2^* - \frac{\alpha}{\beta_2+\alpha}z))} = c/y. \tag{14}$$

Equations (13) and (14) are the second-period analogues of equations (3) and (4). It is easy to check that (13) and (14) admit a solution if and only if z is sufficiently high.[28]

7.2.3 Best-Response Behavior and Beliefs in the Second Period

Next, we explore best-response behavior of agents given their beliefs in the second period. Once again, note that we observe each agent's reported belief about the size of the attack in the second period.

[28] For a special case where the cost of attacking equals 1/2, "high z" means $z > 1/2$. Under this condition, $\theta_1^*(z) < 1/2$, because $\theta_1^*(z)$ is monotonically decreasing in z and $\theta_1^*(1/2) = 1/2$. (See Appendix A for proof.)

Definition 6 *Let* $m(x_2; \widehat{x}_2, \widehat{x}_1)$ *denote the belief about the measure of other agents attacking in the second period for an agent with signal* x_2 *who thinks that others attacked the status quo in the first period if and only if* $x_1 \leq \widehat{x}_1$*, and who thinks that the others attack in the second period if and only if* $x_2 \leq \widehat{x}_2$*.*

This belief is given by the expected size of the attack:

$$m(x_2; \widehat{x}_2, \widehat{x}_1) = E[A(\theta; \widehat{x}_2) \geq \theta | x_2] \tag{15}$$

Note that, since $A(\theta; \widehat{x}_2) = \Phi\left(\sqrt{\beta_2}(\widehat{x}_2 - \theta)\right)$, the belief about the size of the attack $m(x_2; \widehat{x}_2, \widehat{x}_1)$ is decreasing in x_2 and increasing in $\widehat{x}_2$. This result is summarized in Prediction 6, section 2.3.

8 Appendix B: Tables and Figures

8.1 Descriptive Statistics

Table B1 provides descriptive statistics by treatment. The data are pooled across groups and sessions.

Table B1: Descriptive Statistics.

	Sessions 1-4 Cost 20	Sessions 1-4 Cost 50	Sessions 5-6 No New Info	Sessions 5-6 New Info
Min θ	-73.83	-92.73	-107.48	-49.06
Max θ	208.46	188.30	256.30	256.68
Mean θ	73.48	64.40	77.78	74.10
Min x (Stage 1)	-82.55	-106.92	-128.23	-71.84
Max x (Stage 1)	220.85	201.84	277.64	281.59
Mean x (Stage 1)	73.37	64.45	77.80	74.25
Mean # Attackers (Stage 1)	9.16	8.89	6.88	7.18
Mean # Attackers (Stage 2)	1.28	0.89	0.50	1.94
Mean Belief (Stage 1)	8.57	8.35	6.62	7.03
Mean Belief (Stage 2)	1.76	1.58	1.08	2.71
% Successful Attacks (Stage 1)	57%	54%	43%	41%
% Successful Attacks (Stage 2)	1.9%	0%	0%	2.1%
Median Comfort Level with Stats and Probability (out of 5)	4	4	4	4
Number of subjects	120	120	60	60

8.2 Robustness Checks with Additional Regression Specifications

Table B2 presents the results additional regression specifications on first-stage data that include additional controls. We test for the potential effects of the time trend (variable round) and the order of treatments. Specifications (1) and (3) include only data for sessions 1-4, because they include the effects of the order of cost treatments. We find that in the sessions with the low cost rounds followed by the 20 high cost rounds, the probability of attacking is significantly lower than in the rounds with the high cost treatment coming first. On the other hand, the belief about the size of the attack in the first stage in higher in the sessions with the low cost rounds followed by the high cost rounds. Specifications (2) and (4) include only data for sessions 5 and 6, because they include the effects of the order of new information treatments. We find that in the sessions with the no new information rounds followed by the rounds with new information, the probability of attacking and the belief about the size of the attack are significantly lower than in the rounds with the new information rounds coming first. Finally, Specification (2) shows that the probability of attacking decreases in the later rounds of the session, but the effect is only significant at the 10 percent confidence level and is small in magnitude. This implies that the role of learning across rounds does not play an important role for attacking behavior.

Table B2: Determinants of Actions and Beliefs in Stage 1,
Additional Specifications

Variable	Dependent Variables:			
	Action in Stage 1		Belief in Stage 1	
	(1)	(2)	(3)	(4)
Private signal, x	-0.001^{***}	-0.001^{***}	-0.013^{***}	-0.018^{**}
	(0.0002)	(0.0006)	(0.002)	(0.006)
Cost of Attacking	0.0002		-0.004	
	(0.0003)		(0.004)	
Belief about Size of Attack	0.067^{***}	0.068^{***}		
	(0.0015)	(0.0048)		
Actual Size of Attack			0.733^{***}	0.676^{***}
			(0.018)	(0.060)
Round (Time Trend)	0.00003	-0.004^{*}	0.0056	0.038^{**}
	(0.001)	(0.002)	(0.011)	(0.011)
Cost Treatment Order	-0.246^{***}		0.592^{***}	
(1 if Cost 20 first)	(0.002)		(0.061)	
New Info Treatment Order		-0.299^{***}		-0.711^{**}
(1 if No New Info first)		(0.020)		(0.137)
R^2	0.84	0.85	0.80	0.81
Subject Fixed Effects	Yes	Yes	Yes	Yes
Round Fixed Effects	No	No	No	No
No. of observations	4800	1200	4800	1200

Notes: Robust standard errors in parentheses (clustered at the group level). For sessions 5 and 6, only the no-new-information data are used. Significance levels: * 10%, **5%, ***1%.

Table B3 documents that the continuation of the experiment into the second stage reduces the believed number of other attackers by 3 (out of 14).

Table B3: The Impact of a Failed Attack in Stage 1 on Beliefs about the Number of Attackers, No New Information

Variable	Dependent Variable: Belief (OLS)			
	1	2	3	4
Private signal, x	-0.079^{***}	-0.079^{***}	-0.074^{***}	-0.074^{***}
	(0.004)	(0.004)	(0.004)	(0.005)
Cost of Attacking	-0.033^{**}	-0.032^{*}	-0.035^{**}	-0.035^{*}
	(0.011)	(0.016)	(0.013)	(0.020)
Stage 2 Dummy	-3.039^{***}	-2.858^{***}	-3.101^{***}	-2.911^{***}
	(0.332)	(0.356)	(0.342)	(0.369)
1st 10 Rnds Dummy			0.128	0.346
			(0.644)	(0.678)
x × 1st 10 Rnds			-0.007	-0.009^{*}
			(0.005)	(0.004)
Cost × 1st 10 Rnds			0.006	0.005
			(0.015)	(0.015)
R^2	0.63	0.67	0.62	0.66
Subject Fixed Effects	No	Yes	No	Yes
Round Fixed Effects	Yes	Yes	No	No
No. of observations	8820	8820	8820	8820

Notes: Robust standard errors in parentheses (clustered at the group level). For sessions 5 and 6, only the no-new-information data are used. Significance levels: * 10%, **5%, ***1%.

Table B4 reports the additional regression specifications that control for order effects on action and beliefs in the second stage.

Table B4: Determinants of Actions and Beliefs in Stage 2, Additional Specifications

Variable	Dependent Variables:	
	Action in Stage 2	Belief in Stage 2
	(1)	(2)
Private signal, x	0.00003	-0.018^{***}
	(0.0002)	(0.003)
New Information	0.010	0.181
(1 if New Information)	(0.019)	(0.158)
Belief about Size of Attack	0.054^{***}	
	(0.007)	
Actual Size of Attack		0.924^{***}
		(0.101)
Round (Time Trend)	-0.0005	-0.012
	(0.0008)	(0.011)
New Info Treatment Order	-0.008	0.118
(1 if No New Info first)	(0.027)	(0.361)
R^2	0.56	0.54
Subject Fixed Effects	Yes	Yes
Round Fixed Effects	No	No
No. of observations	1395	1395

Notes: Robust standard errors in parentheses (clustered at the group level).
Sessions 5 and 6. Significance levels: * 10%, **5%, ***1%.

Table B5 reports the results of an OLS regression of the expected size of the attack on the private signal and the new information dummy.

Table B5: The Impact of New Information on Individuals' Beliefs about the Number of Attackers in Stage 2

Variable	Belief in Stage 2 (OLS)
Private signal, x	-0.038^{***}
	(0.005)
New Information Dummy	1.358^{**}
	(0.676)
Subject Fixed Effects	Yes
Round Fixed Effects	Yes
No. of observations	1395

Notes: Robust standard errors in parentheses (clustered at the group level).
Significance levels: * 10%, **5%, ***1%.

References

[1] Angeletos, George-Marios, Hellwig, Christian, and Alessandro Pavan. 2007. "Dynamic Global Games of Regime Change: Learning, Multiplicity, and Timing of Attacks." *Econometrica* 75(3): 711-756.

[2] Angeletos, George-Marios and Ivan Werning. 2006. "Crises and Prices: Information Aggregation, Multiplicity, and Volatility." *American Economic Review* 96(5): 1720-1737.

[3] Arifovic, Jasmina and Michael Maschek. 2006. "Currency Crisis: Evolution of Beliefs, Experiments with Human Subjects and Real World Data." Revisions done for *The Journal of Economic Behavior and Organization*.

[4] Atkeson, Andrew. 2000. "Discussion of Morris and Shin's 'Rethinking Multiple Equilibria in Macroeconomic Modelling'." National Bureau of Economic Research *Macroeconomics Annual*.

[5] Baltagi, Badi H. 1995. *Econometric Analysis of Panel Data*. John Wiley & Sons.

[6] Boston, Robert. 2003. "Contagion Without Borders: The Lessons Taught by the 'Tequila Effect' or the 'Brazilian Sneeze'." 2003 Global Investment Conference.

[7] Brunnermeier, Markus K. and John Morgan. 2006. "Clock Games: Theory and Experiments." Manuscript, Haas School, UC Berkeley.

[8] Cabrales, Antonio, Nagel, Rosemarie, and Roc Armenter. 2003. "Equilibrium Selection Though Incomplete Information in Coordination Games: An Experimental Study." Discussion Paper No. 601, Universitat Pompeu Fabra, Barcelona.

[9] Carlsson, Hans and Eric van Damme. 1993a. "Global Games and Equilibrium Selection." *Econometrica* 61(5): 989-1018.

[10] Carlsson, Hans and Eric van Damme. 1993b. "Equilibrium Selection in Stag Hunt Games" in *Frontiers of Game Theory*, ed. by K. Binmore, A. Kirman, and P. Tani. Cambridge, MA: MIT Press, 237-253.

[11] Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47: 225-238.

[12] Chamley, Christophe. 1999. "Coordinating Regime Switches." *The Quarterly Journal of Economics*, 114(3): 869-905.

[13] Chen, Qi, Goldstein, Itay and Jiang, Wei. 2009. "Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows." *Journal of Financial Economics*, forthcoming.

[14] Cheung, Yin-Wong and Daniel Friedman. 2009. "Speculative Attacks: A Laboratory Study in Continuous Time." *Journal of International Money and Finance,* Elsevier, 28(6): 1064-1082.

[15] Cooper, Russell W., DeJong, Douglas V., Forsythe, Robert, and Ross, Thomas W. 1990. "Selection Criteria in Coordination Games: Some Experimental Results." *American Economic Review* 80(1): 218 – 233.

[16] Cooper, Russell W., DeJong, Douglas V., Forsythe, Robert, and Thomas W. Ross. 1992. "Communication in Coordination Games." *Quarterly Journal of Economics* 107(2): 739-771.

[17] Cornand, Camille. 2006. "Speculative Attacks and Information Structure: An Experimental Study." *Review of International Economics* 14(5): 797-817.

[18] Corsetti, Giancarlo, Guimaraes, Bernardo, and Nouriel Roubini. 2003. "International Lending of Last Resort and Moral Hazard: A Model of IMF's Catalytic Finance." NBER Working Papers 10125, National Bureau of Economic Research, Inc.

[19] Costain, James S., Heinemann, Frank, and Peter Ockenfels. 2007. "Multiple Outcomes of Speculative Behavior in Theory and in the Laboratory." Bank of Spain working paper.

[20] Danielsson, Jón and Francisco Peñaranda. 2007. "On the Impact of Fundamentals, Liquidity and Coordination on Market Stability." Economics Working Papers 1003, Department of Economics and Business, Universitat Pompeu Fabra.

[21] Edmond, Chris. 2003. "Information and the Limits to Autocracy." Working paper, UCLA.

[22] Fehr, Ernst and Jean-Robert Tyran. 2005. "Individual Irrationality and Aggregate Outcomes." *Journal of Economic Perspectives*, American Economic Association, 19(4): 43-66.

[23] Fehr, Ernst and Jean-Robert Tyran. 2008. "Limited Rationality and Strategic Interaction – The Impact of the Strategic Environment on Nominal Inertia." *Econometrica* 76(2): 353-394.

[24] Urs Fischbacher. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10(2): 171-178.

[25] Kristin Forbes. 2000. "The Asian Flu and Russian Virus: Firm-level Evidence on How Crises are Transmitted Internationally." NBER Working Papers 7807, National Bureau of Economic Research, Inc.

[26] Freedman, Amy L. 2005. "Economic crises and political change: Indonesia, South Korea, and Malaysia." *Asian Affairs: An American Review*, Heldref Publications.

[27] Glick, Reuven and Andrew K. Rose. 1998. "Contagion and Trade: Why Are Currency Crises Regional?" NBER Working Papers W6806, National Bureau of Economic Research, Inc.

[28] Goldstein, Itay, and Ady Pauzner. 2004. "Contagion of Self-Fulfilling Financial Crises due to Diversification of Investment Portfolios." *Journal of Economic Theory*, 119(1): 151-183.

[29] Haltiwanger, John C. and Michael Waldman. 1985. "Rational Expectations and the Limits of Rationality: An Analysis of Heterogeneity." *American Economic Review* 75(3): 326-40.

[30] Haltiwanger, John C. and Michael Waldman. 1989. "Limited Rationality and Strategic Complements: The Implications for Macroeconomics." *Quarterly Journal of Economics* 104(3): 463-83.

[31] Heinemann, Frank, Nagel, Rosemarie, and Peter Ockenfels. 2004. "The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information." *Econometrica* 72(5): 1583-1599.

[32] International Monetary Fund. 2000. "Recovery from the Asian Crisis and the Role of the IMF."

[33] Izmalkov, Sergei and Muhamet Yildiz. 2009. "Investor Sentiments." *American Economic Journal: Microeconomics*, forthcoming.

[34] Kraay, Aart. 2003. "Do high interest rates defend currencies during speculative attacks?" *Journal of International Economics*, 59(2): 297-321.

[35] Krugman, Paul R. 1979. "A Model of Balance of Payments Crises." *Journal of Money, Credit, and Banking,* 11: 311 – 325.

[36] Meese, Richard and Kenneth S. Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies. Do They Fit out of Sample?" *Journal of International Economics,* 14: 3-24.

[37] Mehta, Judith, Starmer, Chris, and Robert Sugden. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games." *American Economic Review,* 84 (3): 658-673.

[38] Morris, Stephen and Hyun Song Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* 88(3): 587 – 597.

[39] Morris, Stephen and Hyun Song Shin. 2004. "Coordination Risk and the Price of Debt." *European Economic Review* 48: 133-153.

[40] Obstfeld, Maurice. 1996. "Models of Currency Crises with Self-Fulfilling Features." *European Economic Review* 40(3-5): 1037 – 1047.

[41] Jack Ochs. 1990. "The Coordination Problem in Decentralized Markets: An Experiment." *Quarterly Journal of Economics* 105(2): 545-559.

[42] Prati, Alessandro and Massimo Sbracia. 2002. "Currency Crisis and Uncertainty about Fundamentals." Temi di discussione (Economic working papers) 446, Bank of Italy, Economic Research Department.

[43] Rochet, Jean-Charles and Xavier Vives. 2004. "Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?" IDEI Working Papers 294, Institut d'Économie Industrielle (IDEI), Toulouse.

[44] Rose, Andrew K. 2001. "Do We Really Know that Currency Crises are Macroeconomic?" Conférence Internationale "Financial Crises and Recovery", CESifo, Venise.

[45] Schotter, Andrew and Tanju Yorulmazer. 2009. "On the Dynamics and Severity of Bank Runs: An Experimental Study." *Journal of Financial Intermediation*, 18(2): 217-241.

[46] Van Huyck, John B., Battalio, Raymond C., and Richard O. Beil. 1990. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review* 80(1): 234-48.