American Economic Review: Papers & Proceedings 2016, 106(5): 592–596
http://dx.doi.org/10.1257/aer.p20161047

Bayesian Persuasion in Coordination Games[†]

By Itay Goldstein and Chong Huang*

Regime-change games have been studied widely in the economic literature in the contexts of bank runs, currency attacks, political revolutions, and others. Usually, agents need to coordinate an attack against the status quo: once sufficiently many agents attack, the regime changes and agents who were attacking benefit from their attack; otherwise, the regime stays and agents who were attacking it have to bear a cost.

An important angle of coordination games is the transmission of information from the policymaker, or the defender of the status quo, to the agents who have to make a decision whether to attack the regime. The policymaker can attempt to change the information available to agents to increase the likelihood of the survival of the regime. In this paper, we present a model where the policymaker attempts to transmit information in this spirit. Specifically, we study a model where the policymaker can commit to abandon the regime automatically (before agents choose whether to attack or not) if the fundamentals determining the strength of the regime are below a certain threshold. The benefit of committing to abandon the regime more often is that, conditional on the regime not being automatically abandoned, agents update their information to account for the fact that the fundamentals are above the threshold and so their inclination to attack decreases. The disadvantage is that it directly increases the likelihood that the regime will not survive (since it is automatically abandoned).

*Goldstein: University of Pennsylvania Wharton School, 3620 Locust Walk, Philadelphia, PA 19104 (e-mail: itayg@wharton.upenn.edu); Huang: Paul Merage School of Business, University of California, Irvine, 4293 Pereira Drive, Irvine, CA 92697 (e-mail: chong.h@uci.edu). We are grateful to Laura Veldkamp and participants at the 2016 ASSA Meetings for helpful comments and suggestions.

[†]Go to http://dx.doi.org/10.1257/aer.p20161047 to visit the article page for additional materials and author disclosure statement(s).

Analyzing this problem we show that the policymaker finds it optimal to commit to abandon the regime often enough such that when agents see that the policymaker did not abandon it they choose not to attack. Hence, the policymaker chooses to maximize the benefit out of the informational transmission channel from the committed policy. At the same time, the policymaker has no additional benefit in committing to abandon the regime beyond that point as it will cause too frequent abandonment without additional information benefits. The commitment ability of the policymaker is key here. We show that without commitment ability, the policymaker ends up abandoning automatically less often, but this triggers more attacks, such that overall the regime survives less often.

The analysis has implications for various settings. Consider the currency attack model studied in Morris and Shin (1998), where a central bank tries to maintain a fixed exchange rate regime, but is subject to speculative attacks from traders in the foreign exchange market. If the central bank commits ex ante to abandon the regime whenever the fundamentals are sufficiently weak, then seeing the regime being maintained, speculators update positively and the attack becomes less likely. Similarly, in the context of bank runs, as in Goldstein and Pauzner (2005), the bank may like to design a policy whereby investments are liquidated if their fundamentals are relatively weak. The bank may choose to do so more often than is ex post desirable for the purpose of deterring runs when fundamentals are even stronger.

Our paper contributes to the vast literature of coordination games, and in particular global games that are used in our model. This literature goes back to Carlsson and van Damme (1993) and has been applied and extended in contexts like currency attacks (Morris and Shin 1998), bank runs (Goldstein and Pauzner 2005), and others. Specifically, our paper is related to other papers that study endogenous information in global games, such as Angeletos, Hellwig,

and Pavan (2007) and Huang (2015). In these papers, agents get information endogenously from the failures of previous attacks. In our paper, the endogenous information is provided by the policymaker who designs it strategically to affect the likelihood of failure.

Other papers have analyzed the attempts of policymakers to affect the information available to agents in global games models. For example, Angeletos, Hellwig, and Pavan (2006) analyze the central bank's interest rate policy as a tool to signal the strength of the currency, and Edmond (2013) studies how a dictator can manipulate the private information available to agents considering a revolution. However, these policies are costly, which reduces the policymaker's incentives to employ them. The policy we study here is much simpler, as it is just about direct information transmission. In equilibrium, it ends up having no cost, as the policymaker abandons the regime in cases in which it would have been abandoned anyway following an attack. Unlike in these other papers, however, our policy requires commitment on the side of the policymaker.

The design of information in our paper is in the spirit of the growing literature of Bayesian persuasion following Kamenica and Gentzkow (2011). The automatic regime change policy can be viewed as an information transmission mechanism, which generates two "straightforward signals": abandoning the status quo automatically corresponds to a recommendation of attacking, and maintaining the status quo corresponds to a recommendation of not attacking. Our paper contributes to this literature by analyzing the information design in a regime change game, which features coordination among multiple receivers who have heterogeneous private information.[1]

[1] A recent paper by Goldstein and Huang (2016) also analyzes information design when receivers have coordination incentives and heterogeneous information. Their model deals with credit rating policies and how they affect investors' decisions to invest in the firm's bonds and the firm's investment decisions. The feedback effect to the firm's investment leads to a much more complex analysis with very different equilibrium outcomes.

I. A Regime Change Game with Bayesian Persuasion

There are three dates, $t = 0, 1, 2$. A regime has two possible outcomes: the status quo ($R = 0$) and the alternative ($R = 1$). The strength of the status quo is described by θ, which is drawn from the prior distribution $\mathcal{N}(\overline{\theta}, \alpha^{-1})$ at date 1. Such a prior distribution is common knowledge at date 0.

As in other regime-change models, there is a continuum of agents who try to trigger the regime change. Agents are uniformly distributed over $[0, 1]$ and indexed by i. At date 2, conditional on the regime still being in place (more on this below), each agent i has to choose between two actions: attack the status quo ($a_i = 1$) or refrain from attacking ($a_i = 0$). Agents make their choices simultaneously. Denote by A the total measure of agents attacking, the regime will change from the status quo to the alternative following the attack, if and only if $A \geq \theta$.

The strength of the status quo, θ, is unknown to the agents when they make a decision whether to attack or not. However, before making a decision at date 2, any agent i observes a private signal $x_i = \theta + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \beta^{-1})$ is independent of θ and independent across agents. As in standard regime-change games, we consider the case that β is sufficiently large.

At the end of date 2, the agents' payoffs realize. If an agent does not attack, he will get the payoff 0 for sure. If an agent attacks, his payoff depends on whether the regime changes: if the status quo is abandoned, the agent will receive a payoff $1 - c$; but if the status quo remains in place, the agent will get a payoff $-c$. Here, $c \in (0, 1)$ is the cost of attacking.

The new ingredient we add to this standard regime-change framework is to allow the policymaker, who is in charge of the regime, to commit at date 0 to an automatic regime-change policy at date 1 (prior to the coordination game described above). Specifically, after observing the realization of θ at date 1, the policymaker will abandon the status quo if and only if the strength of the status quo, θ, is less than or equal to a threshold y. If $\theta > y$, the policymaker will maintain the regime, but then the agents may force abandoning the status quo based on the coordination game at date 2, as described above. For simplicity and when there is no confusion,

we call an automatic regime change policy with the threshold y: "policy y."

The policymaker can commit on y at date 0. She chooses the policy to maximize the overall probability of the survival of the status quo. This policy acts like Bayesian persuasion because the decision of the policymaker whether or not to maintain the regime at date 1 sends a message to the agents about the strength of the regime. Moreover, how informative this signal is depends on the policy y the policymaker commits to.

We are interested in a monotone perfect Bayesian equilibrium. If there are multiple equilibria in the coordination subgame at date 2, we select the one with the largest measure of agents attacking. That is, different from the "sender-preferred perfect Bayesian equilibrium" selected by Kamenica and Gentzkow (2011), we consider the "policymaker least preferred perfect Bayesian equilibrium." So, the policymaker will choose a policy knowing that for any policy she chooses agents will coordinate on the worst possible outcome for her. Otherwise, if we consider the policymaker's preferred equilibrium, the coordination feature of the model directly implies that the policymaker will choose the policy $y = 0$. Essentially, by abandoning the regime whenever attacking is a dominant action, the policymaker eliminates the lower-dominance region in global-games models, giving rise to an equilibrium where no one attacks after seeing that the policymaker chose to maintain the regime. While this is an interesting observation, demonstrating the potential effect of the automatic regime change policy, it is rather trivial. The literature on coordination games often focuses on the bad outcomes, and so we focus here on the least preferred perfect Bayesian equilibrium.

II. Equilibrium Characterization

We now characterize the policy chosen by the policymaker at date 0. We derive it by backward induction. We start by analyzing the coordination game at date 2, following a committed policy y and the status quo surviving at date 1 (i.e., $\theta > y$).

Considering a monotone equilibrium, given the agents' strategy, the measure of agents attacking is decreasing in the strength of the status quo. Therefore, there must be a $\theta^* \geq y$, such that the regime changes if and only if $\theta \leq \theta^*$. Consequently, maximizing the probability of the survival of the status quo is equivalent to minimizing θ^*. The policymaker will thus choose the optimal y such that the regime change threshold θ^* is minimized.

Now, thinking about agents' decisions, we know that the survival of the status quo at date 1 conveys to all agents that $\theta > y$. Agent i with private signal x_i then forms a truncated normal posterior belief about θ, which leads to the posterior belief of a regime change

$$1 - \frac{\Phi\left[\sqrt{\alpha + \beta}\left(\frac{\alpha\overline{\theta} + \beta x_i}{\alpha + \beta} - \theta^*\right)\right]}{\Phi\left[\sqrt{\alpha + \beta}\left(\frac{\alpha\overline{\theta} + \beta x_i}{\alpha + \beta} - y\right)\right]}.$$

Since such a posterior belief of regime change is strictly decreasing in x_i, agent i will attack if and only if $x \leq x^*$. Here, x^* is determined such that an agent who observes x^* is indifferent between attacking and not attacking:

$$(1) \quad 1 - \frac{\Phi\left[\sqrt{\alpha+\beta}\left(\frac{\alpha\overline{\theta}+\beta x^*}{\alpha+\beta} - \theta^*\right)\right]}{\Phi\left[\sqrt{\alpha + \beta}\left(\frac{\alpha\overline{\theta} + \beta x^*}{\alpha + \beta} - y\right)\right]} - c = 0.$$

Given agents' strategy, because the status quo will be abandoned if and only if $\theta \leq \theta^*$, we have

$$(2) \qquad A(\theta^*) = \Phi\left(\sqrt{\beta}(x^* - \theta^*)\right) = \theta^*,$$

where Φ is the cumulative distribution function (cdf) of the standard normal distribution. That is, at θ^*, the measure of agents attacking is just the same as the strength of the status quo.

We solve x^* as a function of θ^* from equation (2) and substitute the latter into equation (1). Then, the left-hand side of equation (1) becomes a function of θ^*. We denote such a function by $H(\theta^*)$. As shown in Angeletos, Hellwig, and Pavan (2007), when $y \in (0, 1)$, $H(\theta^*) = 0$ has at most finitely many solutions. The largest solution corresponds to the policymaker least preferred equilibrium, because the regime change threshold is the highest. In addition, for any θ^*, H is strictly decreasing in y. Figure 1 illustrates that increasing the committed policy from y' to $\hat{y}$ to y'', leads the H function to keep moving down.

Importantly, when the policy is y', there are two solutions to the equation $H(\theta^*) = 0$, and

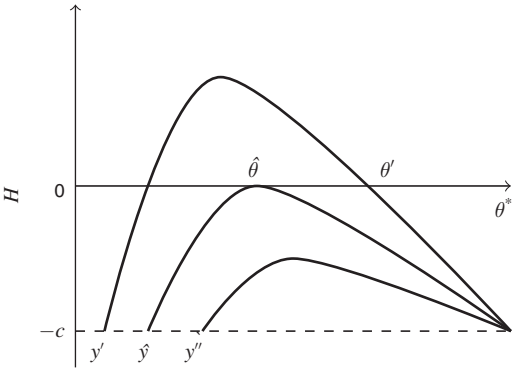

FIGURE 1. *H* FUNCTION

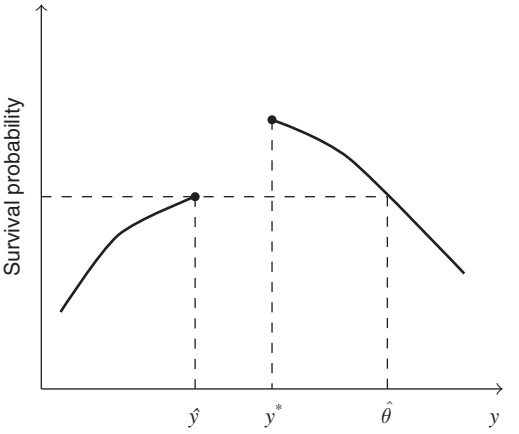

FIGURE 2. SURVIVAL PROBABILITY

the larger solution, denoted by θ', represents the chosen equilibrium in the subgame. Hence, when the policymaker commits to the policy y', the regime change threshold is $\theta' > y'$. When the policy is y'', there is no solution to the equation $H(\theta^*) = 0$, and the subgame has a unique equilibrium in which no agent attacks. Hence, when the policymaker commits to the policy y'', the regime change threshold is just y''. When the policy is $\hat{y}$, the equation $H(\theta^*) = 0$ has a unique solution $\hat{\theta}$. Such a solution corresponds to the equilibrium that is least preferred by the policymaker in the subgame.

Hence, thinking about the threshold below which the regime is abandoned, we can see that it is decreasing in y as y increases to $\hat{y}$, and at that point it makes a discrete jump down, and starts increasing as y continues to increase beyond $\hat{y}$. This suggests that the policymaker would like to set the committed policy just above $\hat{y}$ and as close as possible to it. This creates a technical difficulty, as there is no well defined equilibrium for the whole game: The policymaker will not settle at any point above $\hat{y}$, as she always wants to get closer to $\hat{y}$ (but not set the policy at $\hat{y}$ itself). To overcome this issue, we make a small technical assumption that y cannot be in a small open interval $(\hat{y}, y^*)$. Formally:

ASSUMPTION 1: *The policy y must be in*

$$(-\infty, \hat{y}] \cup [y^*, +\infty),$$

where $\hat{y}$ *is the policy such that* $H(\theta^*) = 0$ *has a unique real solution* $\hat{\theta}$, *and* $y^* \in (\hat{y}, \hat{\theta})$.

With Assumption 1, Figure 2 below illustrates the overall status quo survival probability as a function of y. And the analysis of such a function leads to Proposition 1.

PROPOSITION 1: *Under assumption 1, and under the assumption that the policymaker least-preferred equilibrium is selected at the coordination subgame, the policymaker chooses to commit to the policy* y^* *at date 0, and no agent attacks at date 2 after the policymaker maintains the regime at date 1.*

Intuitively, the policy y^* achieves the optimal balance between the positive informational role of the policy and the negative direct effect it has on the survival likelihood. The threshold y^* is high enough so that in case the regime survives after date 1, agents are optimistic enough about θ that they choose not to attack. As a result, the regime survives the overall game whenever its strength is above y^*. Alternatively, if the policymaker chose to commit on abandoning the regime less often, her maintaining the regime would not reveal enough positive information to speculators and would still trigger attacks such that the overall probability of survival of the regime would drop. At the same time there is no need to increase the committed threshold above y^* as this will lead to abandoning the regime automatically more often without additional informational gain.

III. The Benefit of Commitment

Without commitment, if all agents refrain from attacking provided that the status quo survives at date 1, the policymaker, after observing $\theta \in (0, y^*]$, will deviate to not abandoning the status quo. Because agents cannot detect such a deviation and thus do not attack, the policymaker's deviation is obviously profitable. Hence, the policy in the equilibrium characterized in Proposition 1 is not time-consistent; it requires the policymaker to have the ability to commit. In Proposition 2, we characterize the equilibrium policy when the policymaker cannot commit.

PROPOSITION 2: *When the policymaker does not have commitment ability, any $y \leq \hat{y}$ is an equilibrium policy, but any $y \geq y^*$ cannot be an equilibrium policy.*

We can rank the equilibria when the policymaker cannot commit to a policy by the policymaker's payoff. Since θ^* is strictly decreasing in y when $y \leq \hat{y}$, the policymaker obtains the lowest regime change threshold $\hat{\theta}$ in the equilibrium with the policy $\hat{y}$. Then, because $\hat{\theta} > y^*$, we can see that the policymaker can achieve a higher survival probability if she has the ability to commit. Intuitively, committing to abandon the regime more often than is ex post optimal conveys positive information about the strength of the regime once the regime is maintained and so prevents speculative attacks and limits the likelihood that the regime will ultimately be abandoned.

REFERENCES

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2006. "Signaling in a Global Game: Coordination and Policy Traps." *Journal of Political Economy* 114 (3): 452–84.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2007. "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks." *Econometrica* 75 (3): 711–56.

Carlsson, Hans, and Eric van Damme. 1993. "Global Games and Equilibrium Selection." *Econometrica* 61 (5): 989–1018.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies* 80 (4): 1422–58.

Goldstein, Itay, and Chong Huang. 2016. "Credit Rating Inflation and Firms' Investment Behavior." Unpublished.

Goldstein, Itay, and Ady Pauzner. 2005. "Demand-Deposit Contracts and the Probability of Bank Runs." *Journal of Finance* 60 (3): 1293–1327.

Huang, Chong. 2015. "Defending against Speculative Attacks: Reputation, Learning, and Coordination." Unpublished.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615.

Morris, Stephen, and Hyun Song Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* 88 (3): 587–97.