

Trust and International Organizations *

Songying Fang
Department of Political Science
Rice University
sfang@rice.edu

Randall W. Stone
Department of Political Science
University of Rochester
stnr@mail.rochester.edu

Abstract

Is it possible for international organizations to persuade governments to adopt policy recommendations that are based on private information? If so, under what conditions? In this study we develop a game theoretic model of persuasion that applies to all types of governments, including those that do not face domestic constituency constraints. In our model persuasion takes place on two levels. First, the international institution can send a credible signal about a crisis and prompt the government to take an action in response, and second, it can direct the government's attention to domestic experts and make their expertise policy relevant. The condition under which this effect can take place is that there is a preference difference between the IO and domestic experts, and that the institution holds the more moderate policy position. In such cases, the IO will truthfully reveal its information, thus building trust with the government, and the government will condition its policy on the IO's information. The results suggest that, far from being an obstacle to international cooperation, polarized domestic politics may be a necessary condition for international organizations to exert effective influence.

*We thank Mark Fey, Thomas Dolan, Justin Fox, Jeff Marshall, and seminar participants at the Watson Center for Conflict and Cooperation at the University of Rochester.

Introduction

In recent years the world has witnessed some of the more sobering consequences of globalization. The accelerating pace of climate change, rapidly spreading epidemics and the global financial crisis remind us that a crisis originating in one country has international consequences. Can international organizations serve as credible sources of information in this environment and influence countries' policy responses to crises? The answer seems to be positive: during the outbreak of SARS in 2003, the World Health Organization (WHO) was able to convince the Chinese government to take drastic measures that were widely seen as representing significant breaks with its past practice, including appointing a new Minister of Health who was a health professional rather than a party careerist. At the same time, however, state leaders and informed publics have become increasingly skeptical about whether the instruments of international governance have the countries' best interests in mind when they prescribe policies. The major international organizations, including the international financial institutions, the World Trade Organization and the European Union, repeatedly find themselves in a crisis of legitimacy. The effectiveness of international institutions as information providers depends upon trust. Under what conditions can an international institution acquire credibility in the eyes of a government as a truthful information provider?

The existing literature on international institutions provides important insights into this question with its attention to domestic mechanisms of IO influence (Milner 1997, Dai 2007). However, the literature has so far focused on electoral mechanisms in democracies, so it has little to say, for example, about how an international organization can influence the policies of an authoritarian government. Furthermore, the literature casts international organizations in the most favorable possible light: it assumes that the difference in policy preferences between national leaders and IOs arises because leaders are biased, and IOs can solve the problem by speaking directly to voters who share their preferences and are in a position

to punish the leaders for undesirable policies (Mansfield, Rosendorff, and Milner 2002, Dai 2007, Fang 2008). However, the conflict of interests between states and international organizations may be more fundamental, arising from the fact that IOs are multilateral bodies that necessarily have different policy goals than individual governments. Finally, IOs often need to communicate information that appears to corroborate their known biases. For example, the WHO may be perceived as continually sounding alarms about global health crises, and thus a particular warning by the organization may be taken as nothing more than another manifestation of such a bias. The literature in economics and political science on information transmission focuses on the possibility of communicating information that runs counter to known biases, but generally suggests that communication fails when the true state of the world corresponds to claims that the biased agent has an incentive to make. A general theory of IO crisis management should address all three of these concerns. It should provide a causal mechanism that applies in both democratic and nondemocratic contexts, which sheds light on how IOs can build trust with governments despite their conflicts of interest, and which explains how IOs can communicate information effectively even when the information appears to confirm their known biases. This study is a first attempt at developing such a theory.

Specifically, we develop a game theoretic model of persuasion that applies to all types of governments. In democratic and nondemocratic political systems alike, leaders must decide how to allocate their attention across a variety of issues, and as a result they must be rationally ignorant — or at least, relatively uninformed — about some public policy issues. When an international institution perceives a potential crisis situation, it can raise the profile of one of these issues by sending a signal that the state of the world urgently demands policy change. In response, the government can either choose a policy by itself or consult its domestic sources of expertise. We assume that the leader generally finds it costly to adjust policy, while the IO and the domestic experts do not internalize such costs, but consider their

best policy responses given the state of the world. When the government delegates, therefore, it risks allowing agents with different preferences to influence its decisions in ways that it might not choose were it more fully informed. However, it may nevertheless be preferable to consult domestic experts if policy change is not too costly and domestic experts have valuable expertise. We allow for the possibility that the IO has an incentive to misrepresent its information in order to induce the government to consult the expert. This model thus identifies a conflict of interest between a government and an international organization, but finds conditions under which the international organization can nevertheless effectively transmit information and influence a government's policy in a crisis situation by building trust.

We contrast the results of three versions of the model. In the baseline model there are only unbiased experts, who share the objectives of the IO and want to adopt the most appropriate policy given the state of the world. In the extended model, we allow for the possibility that the domestic expert is biased, with a preference to either take precautions against potential crises even when the objective situation makes them unwarranted, or do nothing at all regardless of the state of the world. In equilibrium we find, in the baseline model where experts are not biased, that there is no truth-telling equilibrium in which the IO honestly reports its private information to the government and the government acts on it.¹ When the cost of implementing the policy is too high or too low, the government's policy choice will not be affected by the institution's signal; when the cost is in the middle range where the institution's information matters, however, the institution has an incentive to misrepresent

¹A truth-telling equilibrium requires that it is incentive compatible for the agent to send a truthful report if the report will be believed and affect the recipient's strategy. When such an equilibrium does not exist, the agent (in this case, the IO) has an incentive to misrepresent its signal. In our model, the IO does not make a false claim about the state of the world, because it does not know the true state of the world with certainty. Instead, when it receives a low signal, it recommends a policy that the government would only prefer to take if the IO had in fact received a high signal.

the signal that it receives in order to motivate the government to consult the expert. As a result, a truth-telling equilibrium does not exist. Ironically, however, the existence of biased experts who always prefer the precautionary policy makes credible communication possible. When experts are unbiased, the IO categorically prefers that the government delegate to the expert, so its advice is not credible. When experts are expected to be biased in favor of an extreme policy with a high probability, however, there exist conditions under which the IO might prefer that the experts control the policy only if it honestly believes that the state of the world is favorable for an activist response. Since the IO sends different messages when it receives different signals in such cases, it is possible for the government to believe the IO's recommendation, and this can lead it to delegate authority to an expert at the suggestion of the institution. Finally, in the case where the expert is biased toward inaction, we again obtain the result that there is no truth-telling equilibrium.

The study builds on the literature that investigates domestic conditions in democracies for institutional influence, but identifies a common mechanism – domestic experts – that can operate in both democratic and nondemocratic contexts. The study also speaks more broadly to the literature on principal-agent relationships. In particular, a significant literature has developed to explain why principals delegate and whether, given divergent preferences, truthful information can be transmitted between the agent and principal (Aghion and Tirole 1997, Bendor and Meiorowitz 2004, Crawford and Sobel 1982, Dessein 2002, Gailmard 2009, Gilligan and Krehbiel 1988). We find that the existence of two agents with divergent interests—the international organization and the domestic policy expert—makes credible information transmission possible in such circumstances. Our ironic result is that the necessary condition for trust is mistrust. That is, the international organization can only credibly persuade a government to delegate authority to its domestic experts if both the IO and the government are known to believe that those experts are biased with sufficiently high probability.

Trust and Information Transmission

International institutions influence state policies and promote multilateral cooperation in a number of ways, but most accounts of their influence deemphasize overt coercion and emphasize instead the role of information (Keohane 1984). Most institutions are incapable of coercing their members, and instead foster cooperation through providing information and expertise, or authoritatively defining rules and expectations (Barnett and Finnemore 2004, Reinhardt 2001, Davis 2003). Persuasion is thus key to the influence of international organizations.

Trust is a necessary condition for successful persuasion. If governments perceive that international institutions are pursuing policies detrimental to their interests, they disregard their policy advice. Cheap talk is cheap; in general, it is difficult to credibly share information when there is a conflict of interest (Fearon 1994). Recent contributions to international relations have treated trust as a fundamental factor in international security. Sartori (2005) argues that the reputation for truth telling plays an important role in avoiding international conflict, and Kydd (2005) explores the effects of trust and mistrust on conflict and cooperation during the Cold War. Trust is similarly important in the relations between states and IOs, and is often problematic because international organizations have incentives to misrepresent their private information so that their preferred policies will be chosen by governments. Prominent accounts of the International Monetary Fund, for example, focus on the tension between national objectives and international agendas (Woods 2006). Therefore, potential conflict of interest between international organizations and national governments is an essential feature of a model of persuasion, and the question is whether truthful information can be transmitted in this environment.

The economic literature on information transmission between a principal and an agent finds that the information content of messages declines as interests become more conflictual

(Crawford and Sobell 1982). An associated literature on third party mediation in political science finds that biased mediators are more effective than unbiased ones, because biased mediators can credibly convey information that is contrary to their own biases (Calvert 1985, Kydd 2003). Similarly, a literature in economics explores the conditions under which governments can credibly convince voters that the state of the world justifies policy change, and concludes that such arguments are only credible when they run counter to the government's known policy preferences (Cukierman and Tommasi 1998, Drazen and Masson 1994). All of this literature suggests that persuasion should be difficult for international organizations, because conflicts of interest frequently arise between international organizations and the governments that they attempt to coordinate. Moreover, the problem may be intractable in crisis situations, because in those cases international organizations have to convince skeptical governments to heed their alarms, in spite of the fact that they have incentives to send warnings even in good times. In other words, in order for IOs to be effective information providers, they need to gain trust even when their information seems to confirm their biases. Under what circumstances can persuasion succeed in spite of the fact that governments know that international institutions have different objectives than they do?

While the existing formal literature on international organizations recognizes that there is often a conflict of interest between a government and an IO, it sidesteps the problem of trust that is our focus by assuming that preferences diverge because national leaders are biased. The standard set-up for such models assumes that IOs have the appropriate policy recommendations and even share the interest of the public, but that certain types of leaders have incentives to adopt short-term or self-interested policies. The conflict of interest is thus characterized as one existing only between neutral IOs and biased leaders (Dai 2005, 2007, Fang 2008, Mansfield, Rosendorff, and Milner 2002). This interpretation often works in the contexts of international trade and conflict; however, in a global financial or environmental crisis, it is more likely that preferences diverge simply because IOs are multilateral bodies. In

the best case scenario, the preferences of international institutions reflect some aggregation of the preferences of their member states. In a crisis situation where global public health or economic stability are in the balance, an institution is unlikely to have the same perspective as the government that is in charge on the ground: the local government does not necessarily internalize the global consequences of the crisis; at the same time, the government, rather than the IO, bears the costs of the policy responses. The conflict of interest thus runs deeper in a crisis situation, and cannot be escaped by replacing a biased leader or relying on domestic constituencies to constrain her.

Moreover, the research agenda that seeks domestic mechanisms to explain IO influence has so far focused on democracies and democratizing countries. These studies typically assume that the international organization has superior information about some state of the world that affects the desirability of a policy, and its decisions send signals to the domestic electorate, which is in a position to punish leaders who choose undesirable policies (Chapman 2009, Fang 2008). Yet empirically, we observe that autocracies pursue membership in prominent international institutions and play a significant role in them. For example, China is becoming increasingly integrated into the system of governance by international organizations. It has learned to use the dispute settlement mechanism of the WTO to its advantage, launching more complaints in the last year than any other country. It has taken a leading role in driving reform of the distribution of vote shares in the IMF and World Bank, and has contributed \$50 billion to the IMF's New Arrangements to Borrow. Influence does not flow in just one direction, either. The extensive engagement has been transformative for China as well as important for international organizations. China's efforts to be seen as playing a constructive role in international regimes for issues as diverse as climate change, health and trade have shifted the balance of power within the government, opened up space for new policy initiatives and subtly reshaped China's foreign and domestic policies. Electoral control does not explain such changes in behavior by nondemocratic countries, which sug-

gests that other mechanisms are operative. This deficiency in the existing literature reflects in part the difficulty in identifying a causal mechanism for nondemocratic regimes, and is particularly problematic in the context of global crises, which affect democratic and authoritarian countries alike. This suggests the need to find mechanisms that apply more generally in democratic and nondemocratic contexts.

The outbreak of Severe Acute Respiratory Syndrome (SARS) in China in 2003 serves as an illustration of the mechanism in our model. The new administration of president Hu Jintao and premier Wen Jiabao came into office in March in the midst of a rapidly spreading epidemic, but initially the central government was deeply uncertain about the severity of the crisis. Lower-level officials had much better information than the central government, because censorship, hierarchical reporting structures and career incentives filtered out bad news before it could reach the top. There was a conflict of interest between experts and government officials: health experts favored publicity to minimize fatalities regardless of the severity of the crisis, while party functionaries preferred to deny the severity of the crisis to minimize social unrest. Indeed, provincial officials in Guangdong sought to hide the outbreak in January and forbade newspaper reports in February, and as late as March the Ministry of Health prohibited Beijing hospitals from disclosing infections. The World Health Organization played a pivotal role by drawing international attention to the epidemic on March 15, which convinced the leaders of the central government to intervene. The new leaders carried out a purge of top officials, including the Minister of Health, and in a break with past practice, the new minister was a health professional rather than a party careerist.² An international organization had exerted persuasion over an authoritarian government, which led to personnel changes that spelled a lasting shift in health policy.

Combining domestic experts and international organizations provides a mechanism to

²“Outbreak Gave China’s Hu an Opening; President Responded to Pressure Inside and Outside Country on SARS,” *The Washington Post* (May 13, 2003, Tuesday): A1.

explain the spread of ideas and change in state behavior.³ IOs influence government policy by engaging in persuasion, which in terms of our model means that rational government agents choose (or may choose) to adjust their policies in response to signals sent by the international organization. Persuasion is only possible if the government trusts the IO as a source of information, and in the context of the model, trust is only possible if truth-telling is an optimal strategy for the IO. Therefore, we solve for truth-telling equilibria in which the IO truthfully reveals its signal about the state of the world, and in turn, the government believes the message sent by the IO. We find that domestic experts play a key role in allowing effective information transmission to take place at the international level. The analysis thus identifies *conditions* under which trust can be built between an IO and a government, which allow the IO to influence the government's policy through its message. Such theoretical analysis that applies to all types of governments does not exist in the literature, and the findings will have important implications for institutional reforms and global crisis management.

In our discussion, we have referred to delegation to a domestic expert and consulting a domestic expert interchangeably, because we have analyzed a version of the model presented below that required the government to delegate to an expert agent in order to benefit from the expert's information and an alternative version that allowed the government to request a report from the expert and retain the decision of whether to implement the advice. In the context of our model, the results of the two variants are identical: whenever a government would choose to delegate to an expert it would also choose to consult with an expert and implement the expert's advice, and vice versa. Thus, our results are robust to alternative assumptions about the extensive form of the game. It is helpful for substantive interpretation of our results that the results hold both for delegation and consultation, because in some

³Domestic experts can be thought of as policy entrepreneurs, which figure prominently in work in international relations and comparative politics on the influence of ideas (Hall 1988, Goldstein and Keohane 1993). For example, Checkel (2009) describes the origin of the New Thinking under Gorbachev in terms of policy entrepreneurs.

substantive contexts credible consultation requires delegation, and in others governments may be reluctant to consult experts, but both conditions may not hold in all contexts.⁴

Model

There are three actors in our model: an international institution, I , a government, G , and a domestic expert, E . Suppose the state of the world is $\theta \in \{0, 1\}$, where 0 represents no crisis, and 1 represents a crisis situation. Each state occurs with equal probability. The institution receives a private signal about the state of the world, $s_I \in \{0, 1\}$, and then sends a publicly observable message, $m_I \in \{0, 1\}$, to the government about the signal. If the institution is truthful, the message corresponds to the signal; if the institution misrepresents the information, the message is the opposite of the signal. Assume $Pr(s_I = 0|\theta = 0) = Pr(s_I = 1|\theta = 1) = \alpha$, where $\alpha > 1/2$. That is, the institution's signal is informative, so that it improves on the prior of the government, but is not perfectly correlated with the true state of the world. After receiving the institution's message, the government can choose a policy or choose to consult the domestic expert. The expert receives her own private signal, $s_E \in \{0, 1\}$ about θ . Given the state of the world, the expert's signal is independent of the institution's, and is of higher quality. That is, assume $Pr(s_E = 0|\theta = 0) = Pr(s_E = 1|\theta = 1) = \beta$, where $\beta > \alpha$. The quality of each signal is common knowledge.

⁴Without delegation experts may not have incentive to collect information or specialize (Gilligan and Krehbiel 1988, Aghion and Tirole 1997). In addition, Dessein (2002) finds that when the preference difference between the principal and the agent is large, the agent has an incentive not to fully reveal private information as analyzed by Crawford and Sobel (1982), and delegation, which grants authority over the use of critical resources, is preferable to communication. He further shows that delegation with veto-power is only beneficial for the principal for large divergences in preferences. Finally, Gailmard (2009) shows that without credible commitment any mechanism that elicits private information reduces to cheap talk and again delegation is preferable to such mechanisms.

Suppose there are two types of experts: unbiased and biased types. The unbiased type shares the preference of the institution, and wants the government to take the best action given the state of the world. The biased type always prefers a particular policy, and thus always has an incentive to lead the government to take that action. In the baseline model we assume that the expert is unbiased; we then extend the model to cases of biased experts coexisting with unbiased experts.

The sequence of play is as follows. Nature chooses the state of the world, θ . The institution receives s_I and then sends a publicly observable message, m_I , to the government. After receiving the message, the government can either choose a policy, x , at this point and end the game, or delegate the decision-making to a domestic expert. If G delegates the decision to the expert, the expert chooses a policy to implement based on m_I and its own signal s_E . The sequence is depicted in Figure 1.

Now we specify the actors' utility functions. Assume that the institution wants to adopt a policy that matches the state of the world. This preference is captured by the following utility function:

$$u_I = \begin{cases} -x & \text{if } \theta = 0 \\ x - 1 & \text{if } \theta = 1 \end{cases} \quad (1)$$

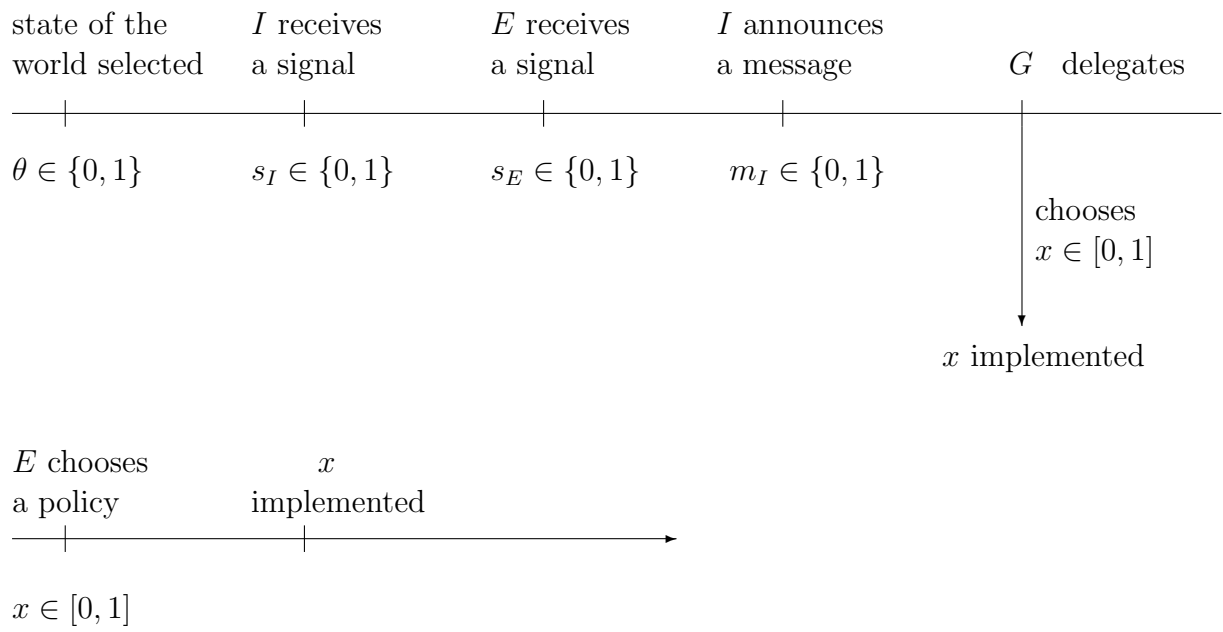
The interpretation of the utility function is straightforward: when the true state of the world is 0, the institution prefers a policy that is as close as possible to 0; when the true state of the world is 1, the institution prefers a policy that is close to 1.

The government wants to take an action that matches the state of the world, but it has an additional concern about the *cost* of implementing a policy, k . We therefore capture its utility function with the following:

$$u_G = \begin{cases} -x - kx & \text{if } \theta = 0 \\ x - 1 - kx & \text{if } \theta = 1 \end{cases} \quad (2)$$

Finally, assume that the utility function of an unbiased expert is identical to that of the

Figure 1: Time Line of the Game



institution. Like the institution, the expert does not internalize the cost of implementing the policy; her only concern is to adopt the most appropriate policy given the state of the world. This creates a tension between her policy preferences and the government's. The model is a game of incomplete information, and it is solved for pure strategy perfect Bayesian equilibria (PBE). We assume that the institution does not play weakly dominated strategies, and when the government is indifferent between implementing its own policy and delegating policy to the expert, it implements its own policy. This is equivalent to assuming that there is a small cost of delegating the policy to the expert.

After analyzing the case where the expert is unbiased with certainty, we extend the model to the case where the expert may be biased toward policy $x = 1$, i.e., with some probability her utility function is $u_E = x - 1$. This scenario captures the selection bias in expert recruitment. That is, experts may be policy outliers: they think that the issues related to their expertise are important, and they self-select into their areas of expertise for that reason. For instance, environmental scientists tend to be environmentalists. We assume that the expert is biased with probability π .

We finally consider the case where the expert may be biased toward policy $x = 0$, i.e., her utility function is $u_E = -x$. It is perhaps more apt to label the expert in this case as a *bureaucrat* who benefits from policy inertia and prefers not to take any action in the face of a crisis. The Chinese party functionaries during the SARS outbreak who preferred to deny the severity of the crisis to minimize social unrest serve as an example. Again, we assume that the expert is biased with probability π .

Equilibrium Results

Our focus is whether the international organization has an incentive to misrepresent the signal that it receives, in which case the government disregards any messages it may send.

In other words, we will be solving for truth-telling equilibria. We present three results, corresponding to our three scenarios regarding the expert. We first analyze the case where the expert is unbiased, which provides a baseline for the other two cases.

Suppose the government believes that the international organization truthfully reveals its signal, and the government decides whether to delegate policy to the expert. In this case, all else equal, the government becomes more willing to delegate policy to the expert as the expert's signal (β) becomes more accurate, but becomes less willing to delegate as the institution's signal (α) becomes more accurate. Intuitively, the more accurate is the expert's signal, the better the policy choice that she makes; therefore, the government has more incentive to consult the expert. However, as the IO's signal becomes more accurate, the government is able to use it to choose better policies without surrendering control, and the marginal policy improvement from delegating policy to the expert is mitigated by the fact that the expert ignores the marginal cost.

The unit cost of implementing a policy, k , plays a critical role in the government's decision about whether to delegate. If the cost is high, the government prefers to choose its own policy to avoid the prospect that the expert chooses a more extreme policy than the government prefers. Specifically, depending on the message sent by the institution, there is a threshold cost, $k_1 = \frac{\beta - \alpha}{\alpha + \beta - \alpha\beta}$ if $m_I = 0$, and $k_2 = \frac{\alpha + \beta - 1}{1 - \alpha - \beta + 2\alpha\beta}$ if $m_I = 1$, above which the government will not delegate the decision-making to an expert.

In contrast, the international organization always prefers delegation to an unbiased expert, regardless of the signal it receives. The expected utility for the IO is always higher when the expert rather than the government chooses policy, because the expert's signal is more accurate than its own, so the expert is positioned to choose a better policy. Furthermore, the government has an incentive to choose a lower policy action than is optimal from the IO's perspective, because the government bears the cost of implementing the policy.

Does this conflict of interest create an incentive for the international organization to misrepresent its private information to the government? Such an incentive exists in some states of the world if truth-telling will result in the government choosing a policy by itself, while sending a warning message will result in the government delegating the decision to the expert. When the cost falls below k_1 or above k_2 , the institution's message cannot change the government's behavior because the government either always delegates ($k < k_1$), or always chooses its own policy ($k > k_2$). As a result, the institution's message can be arbitrary, and thus can be truthful in a trivial sense. If the cost falls between the two thresholds, i.e., $k_1 < k < k_2$, the government will not delegate if it believes that the institution received the signal $s_I = 0$, but will delegate if it believes the institution received $s_I = 1$. Within this range, the institution's signal is sufficiently informative that the expected benefit from letting the expert choose policy outweighs the expected cost of implementing the policy when $s_I = 1$. However, if the government believes the institution's message and acts accordingly, this scenario provides an incentive for the institution to mislead the government. Specifically, by reporting that there is evidence of an imminent threat of crisis when its signal indicates otherwise, the institution could prompt the government to delegate the policy to the expert. Understanding that the institution categorically prefers delegation, the government expects it to send the message $m_I = 1$ regardless of the signal it receives, and consequently disregards the institution's message. As a result, there is no truth-telling equilibrium when $k_1 < k < k_2$. We summarize the result in the following proposition.

Proposition 1. *If all available domestic experts are unbiased, no truth-telling equilibrium exists in which the international organization truthfully reports its signal, and the government conditions its decision whether to delegate on the IO's message.*

The implication of Proposition 1 is that when domestic experts are known to share the preferences of the international organization precisely, and these preferences differ from those of the government, there are no circumstances under which the institution can influence the

government's decision about whether to delegate. The government ignores the institution's message altogether, regardless of the cost of implementing the policy and regardless of how well informed the international organization may be.

We now turn to the case where domestic experts may be biased. Specifically, the expert is biased toward policy $x = 1$ with probability $\pi > 0$, and unbiased otherwise. A biased expert, in this sense, may be thought of as an agent who makes principled decisions rather than pragmatic decisions conditioned by the prevailing circumstances. Both types of expert will choose according to her true preference. This means that the biased expert will choose $x_E = 1$, while the unbiased type will choose a policy that is consistent with her updated belief about the state of the world.

Interesting new findings emerge in this scenario. First, the international organization no longer categorically prefers policy delegation: if π is sufficiently large, i.e., the probability is high that the domestic expert is biased, the IO may prefer that the *government* choose its own policy rather than delegating. Second, the expected utility for the government from delegation decreases in π , so the government is more reluctant to delegate. The threshold cost below which the government will delegate regardless of the IO's message, k'_1 , and the cost threshold below which it will delegate policy to the expert if it believes the IO's message that the state of the world requires policy change, k'_2 , are smaller than k_1 and k_2 in the baseline case, respectively. This means that the government is more likely to choose its own policy if it believes that there exist experts who will always recommend extreme policies.

As in the baseline case, we have three cases to consider to determine whether the institution has an incentive to reveal its information to the government truthfully. When the cost falls below k'_1 or above k'_2 , the institution's message cannot change the government's behavior because the government either always delegates ($k < k'_1$), or always chooses its own policy ($k \geq k'_2$). If the cost falls between the two thresholds, i.e., $k'_1 \leq k < k'_2$, the government will not delegate if the institution reports $m_I = 0$, but will delegate if the institution reports

$m_I = 1$, because the expected benefit from letting the expert choose the policy outweighs the expected cost of implementing the policy when $m_I = 1$. As in the baseline, this scenario makes it possible for the institution to manipulate its message, so we turn next to whether the institution has an incentive to do so in this case.

Indeed, we find that if π is sufficiently small, i.e., the domestic expert is very unlikely to be biased, a truth-telling equilibrium does not exist for $k'_1 < k < k'_2$, as in the baseline. However, if π is large, i.e., the probability is high that the domestic expert is biased, it is not in the institution's interest to prompt the government to delegate. Consequently, the international institution sends truthful information to the government, and the government is able to utilize its message to decide whether or not to delegate. Ironically, biased domestic experts are a necessary condition for an international institution to exert influence. Governments and international institutions make common cause when neither fully trusts the domestic experts. The international institution exerts its influence by empowering domestic experts, but it can most credibly empower policy extremists, rather than the pragmatic, case-by-case decision makers that it and the government would most prefer. We therefore have the following result.

Proposition 2. *If there exist biased experts who always prefer adopting an extreme policy, and the probability of encountering such an expert is sufficiently high, there exists a unique truth-telling equilibrium in which the institution truthfully reveals its information. Consequently, the government uses the international institution's message to decide whether to delegate policy to a domestic expert.*

These two results provide a rather surprising interpretation of the role of domestic experts in facilitating international influence on governmental policies. Contrary to our intuition, the more reliable domestic experts are known to be, the more difficult it is for international institutions to persuade governments to consult them. If domestic experts are known to be pure technocrats who fully share the institution's preferences and prescribe policies strictly

based on their expertise, this provides an incentive for the international institution to misrepresent its private information. As a result, the institution cannot credibly persuade the government to delegate the policy-making decision based on the institution's message. On the other hand, the existence of activist experts who would prescribe extreme policies deters the institution from manipulating its message, making the institution a credible source of information for the government.

A natural question is whether the result from Proposition 2 holds in model 3, where the domestic expert may be biased towards the other extreme policy, $x = 0$.⁵ In this case, the result does not hold for any size of π . When the expert is biased towards policy 0, the international institution again has an incentive to misrepresent its information; consequently, there is no truth-telling equilibrium. The intuition is as follows: when the international institution receives signal 0, it knows that truth-telling will lead the government to do nothing (choosing policy 0). On the other hand, if the institution misrepresents its signal and the government delegates, there is at least some chance that an unbiased expert will be drawn. There is no downside to delegation from the international institution's point of view: the biased expert chooses the same policy as the government ($x = 0$), while the unbiased expert chooses $x = 1$ only if the expert's signal indicates that this is optimal. Since the expert's signal is more accurate than the institution's signal, this is optimal regardless of the signal the institution receives. As a result, the institution always has an incentive to manipulate its message. The following proposition captures this analysis:

Proposition 3. *If experts have only a status-quo bias, no truth-telling equilibrium exists in which the international institution truthfully reports its signal and the government conditions its decision about whether to delegate on the institution's message.*

⁵As discussed earlier, a more appropriate interpretation of the "expert" in this case is a bureaucrat who has a status quo bias.

Discussion

The fundamental result of our analysis is that international organizations can influence government policies and promote delegation to policy experts whose preferences diverge from the government's, but only when there is a sufficiently high probability that the policy experts in question are biased in favor of innovative policies. Bias, in this sense, means that the policy experts categorically prefer policy innovation irrespective of the state of the world, which is what justifies policy change from the points of view of the government and the international organization. Health professionals prefer to take precautions to contain potential epidemics, even if it is possible that the public health benefits may be meager, while Chinese officials were prepared to accept the necessity of publicizing the SARS crisis only if the outbreak proved to be severe. Neoliberal economists prefer market-oriented reform irrespective of the business cycle, but governments might feel that such reforms are justified only if there is an impending financial crisis. Environmental scientists prefer to reduce the degree of global climate change regardless of whether new data indicate that projected global temperatures will be higher or lower than previously believed, but these projections may be critical to the cost-benefit calculations of politicians. The international organization and the domestic expert share a bias in the sense that neither pays the costs of the new policy, so both are suspect from the government's point of view. However, the international organization shares the government's interest in matching the policy to the state of the world; the domestic agent does not if she is biased, and instead categorically prefers policy change.

When domestic experts are known to share the international organization's preferences precisely, its message is uninformative because the institution has an incentive to advise the government to delegate in every state of the world. Consequently, the government cannot learn anything from the international organization that would allow it to update its beliefs, so the IO's message is disregarded. The same is true when domestic experts

may be biased, but only in favor of the status quo: again, the international organization categorically prefers delegation, so its message is uninformative. (In this case, the biased type of domestic expert chooses the same policy as the government, so delegation costs the international organization nothing, while the unbiased type of the domestic agent chooses policy innovation only when its superior information about the state of the world indicates that it is justified.) Ironically, in the cases of unbiased or status-quo biased experts, where the government would be most willing to delegate *ex ante*, the international organization cannot provide credible information, so domestic experts are ignored.

When domestic experts may be biased policy activists, on the other hand, international organizations are able to build trust with governments because neither fully trusts the domestic experts. The possibility that domestic experts may be more radical than the international organization disciplines the IO, deterring it from announcing that the state of the world favors radical reform unless it truly believes that this is the case. Intuitively, international organizations are most credible when they are moderate, relative to the policy alternatives represented in domestic politics. This analysis has implications for the role of domestic experts in the policy making process, for the conditions under which international organizations can build trust, and for the types of policies that international organizations can fruitfully promote.

The first implication of the model is that the influence of international organizations depends on the existence of reform-minded policy communities within the countries they are attempting to influence, and this resonates with recent empirical work. The literature on policy diffusion suggests that international organizations play a catalytic role in spreading policy ideas, but that these ideas take root when conditions are ripe within domestic politics. Recent research on the IMF has emphasized the importance of domestic policy-making elites who form transnational alliances with international bureaucrats and share similar training. Jeffrey Chwieroth (2007) argues that liberalization of the capital account is strongly asso-

ciated with the presence of neoliberal economists in important policymaking positions in the liberalizing countries. In other work, he argues that IMF lending is influenced by the presence of "sympathetic interlocutors," or officials in developing countries that share similar perspectives with IMF officials because they share a common educational background (Chwieroth 2008). Our model proposes a mechanism for this effect that emphasizes the problem of credible communication, and finds that an international organization can generally promote liberalization only when there is a domestic constituency for reform that is believed to hold more extreme views than the IO's.

On the other hand, a second implication of this analysis is that the influence of international organizations should decline as domestic policy experts moderate their positions. The surprising implication of the model is that international organizations are most influential when domestic policy experts are not trusted; when domestic policy experts become credible and move into the mainstream, in contrast, the scope for international organizations to influence policy is curtailed. Continuing to use the example of the IMF, our model suggests that the reason neoliberal policymakers are critical to IMF influence is that they represent policy extremists, whose far-right positions allow the IMF to signal credibly the urgency of policy reform. The IMF attempts to moderate inflation and increase macroeconomic stability, but does not categorically prefer rapid privatization, fixed exchange rates, or shock therapy. When it signals that the economic situation is critical, however, this tips domestic politics in favor of the experts, who may turn out to be radical reformers. This, in fact, is what makes the IMF's signals credible, because it would not prefer extreme economic policies if the economic situation were not critical.

Russia in the 1990s provides examples both of extraordinary IMF influence and of its gradual decline. The new Russian president, Boris Yeltsin, chose Yegor Gaidar, a young economist with no track record, to head his economic team because all of the more experienced candidates were closely associated with central planning. Yeltsin was a particularly

uninformed economic policy maker, and he subsequently vacillated between reform and retrenchment, but he initially gave Gaidar wide latitude to develop policy. A vigorous debate was raging among Russian economists about whether to follow the example of Poland and carry out shock therapy to bring inflation under control, or to implement a more gradual set of economic reforms. The reformers faced an impending political deadline: Gaidar expected his "kamikaze government" to be fired within a year because of the painful measures necessary to stabilize prices, so reform had to be rapid for political reasons regardless of the economic merits of sequencing or pacing (Aslund 1995). In terms of our model, the Gaidar government categorically preferred rapid reform irrespective of the economic fundamentals, which made the IMF a moderate advisor in the Russian context. IMF advice, however, gave the Gaidar team crucial political support, because it convinced Yeltsin that the economic crisis was sufficiently severe to warrant radical reform measures. Gaidar's prediction proved to be accurate, since Yeltsin fired him in order to appease his critics in parliament. His successor as the leading Russian reformer, Anatolii Chubais, similarly promoted a mass privatization plan that was designed with politics more than economics in mind: it was intended to shatter the branch ministries' control of industry and create a constituency for reform. Whether mass privatization was the best way to create effective owners was a secondary consideration, and the IMF was initially skeptical. IMF support was ultimately instrumental in implementing mass privatization, however, and Chubais' partisanship made the IMF a moderate advisor.

Russia finally brought inflation under control in 1995 by using interest rates to target the exchange rate with the dollar. Over the next several years, however, the IMF found Russian reformers resistant to its calls for budget cuts and tax increases, while expanding fiscal deficits undermined the credibility of the ruble's exchange rate band. Political constraints had become intense. Parties representing the far right and the far left won parliamentary elections in 1995. In 1996 Chubais was busy running Yeltsin's reelection campaign for the

presidency, and the government ran up substantial fiscal deficits during the election season. Over the next two years, many reformers called for fiscal austerity, and the IMF signaled its displeasure by suspending loan tranches, but the urgency of Russia's fiscal problems seemed to be diminished by rapid capital inflows that made it possible to refinance government debt (Aleksashenko 1999). In terms of our model, Russia's economic policy establishment had moved towards the center of the political spectrum, and the IMF lost influence because it was no longer perceived as a more moderate voice than the reformers. By the time Yeltsin realized that the state of the world called for drastic policy changes in 1998, this had become obvious to global financial markets as well, and it was too late to halt the slide towards devaluation and default.

The IMF had a similar degree of access in the early 1990s in Argentina and Brazil, both countries that had experimented with populist governments and hyperinflation after their transitions to democracy. In both cases, the IMF succeeded in impressing upon new governments that overcoming hyperinflation required drastic macroeconomic policy corrections, and both countries turned to economists without political experience to head their economic policy teams. Both countries opted for rather extreme forms of exchange-rate based stabilization: Brazil adopted the Real Plan, which pegged its new currency to the dollar, and Argentina adopted its Convertibility Law, which established a currency board regime and fixed the peso at parity with the dollar. The IMF was initially critical of both proposals, arguing that while a fixed exchange rate can be a useful nominal anchor to slow inflation during a transition, it is not a sustainable regime and requires an explicit exit strategy. Again, in terms of our model, the radical proposals of domestic experts rendered the IMF a moderate voice. However, the IMF's insistence on the severity of the crisis and the need for radical policy change lent crucial credibility to the domestic policy teams' calls for economic austerity. By the late 1990s and the first years of the next decade, however, the IMF's leverage had deteriorated markedly. IMF officials repeatedly called on the governments of Brazil

and Argentina to reduce their budget deficits and their rapidly accumulating foreign debts, which were undermining the framework of their exchange rate policies. Budget deficits fueled domestic inflation and created current account deficits, and in Argentina mounting public debt threatened to unleash a spiral of exploding debt dynamics that posed the specter of default (Mussa 2002). The entrenched policy teams in both countries included economists trained in top departments, however, and they resisted the IMF diagnosis that their fiscal problems had reached emergency status (IEO 2003, 2004). Brazil suffered a sharp currency crisis in 1999, and Argentina entered a protracted crisis in 2001 that led to default and economic collapse. As in Russia, the IMF lost influence over time in Brazil and Argentina as domestic reformers moved into the political mainstream.

A third implication of the model is that international organizations maximize their influence if they produce relatively moderate policy proposals. Again, the International Monetary Fund provides a good illustration. For most of its history, the IMF has lagged behind rather than led the call for market liberalization. The IMF was originally intended to safeguard the system of fixed exchange rates foreseen at Bretton Woods, and this mandate was assumed to require the maintenance of capital controls (Helleiner 1994, Pauly 1997). This stance shifted only gradually, and it was not until abolishing capital controls had become the agenda of the European Communities and the OECD that IMF Managing Director Michel Camdessus sought to make it a part of the Fund's official mandate (Abdelal 2007). This initiative was largely discredited by the Asian financial crisis of 1997, and the effort to change the formal rules was abandoned in the following year, although pushing for capital liberalization became the unofficial policy of the IMF in the 1990s. Meanwhile, the IMF carried out a "silent revolution" that dramatically increased the comprehensiveness and intrusiveness of the conditionality attached to its loans (Boughton 2001, Gould 2003, 2006, Dreher and Jensen 2007, Dreher and Vaubel 2004).

The striking influence of the International Monetary Fund in the early 1990s coincided

with revolutionary developments in Eastern Europe that discredited state management of the economy and created political space for experimentation with market-oriented reforms. In the process, the IMF moved beyond its traditional competency in macroeconomic policy and began to grapple with the structural features of national economies presumed to retard growth. In the wake of the Asian crisis, the Fund initially stepped up its conditionality and pioneered a new round of conditions targeted at the financial sector. All of these changes—capital liberalization, structural conditionality and financial sector conditionality—led to a profound crisis of legitimacy for the IMF. Even IMF insiders and sympathetic outsiders began to speak of a problem with legitimacy, and this was generally linked to calls to “streamline” conditionality and increase the degree of “ownership” of programs by member governments (Hills, Peterson and Goldstein 1999, Khan and Sharma 2001, Drazen 2002, IMF 2005). The most visible symptom of the crisis was that by early 2008 the IMF was left virtually without borrowers, and because the Fund’s income comes from the interest on its loans, it was forced to announce a ten percent staff reduction. Those constraints were relaxed by the advent of the global financial crisis in the fall of 2008, which brought the IMF many new customers, but the calls for reforming the IMF and increasing the share of votes on the Executive Board held by developing countries have grown. The implication of our model is that IMF influence was maximized in the early 1990s, when it was relatively moderate compared to the radical reformers taking office in Eastern Europe and Latin America. It subsequently lost influence, in part as a result of its own drift to the right, which made governments skeptical when the Fund tried to pose as an honest broker between themselves and their own domestic experts.

Conclusion

Is it possible for international organizations to persuade governments to delegate policymaking authority to experts? If so, this provides a mechanism by which international organiza-

tions can exert a socializing influence: experts will gradually change policies to bring them into line with international norms. Policy change imposes costs on government elites, however, so the question is why a government would allow itself to be convinced to devolve power to experts. This behavior becomes particularly puzzling when the interests of government elites diverge from the interests of international organizations, so that the arguments that IOs use are immediately suspect. We analyze this situation game theoretically.

Our model is a game of incomplete information in which an international institution and a domestic policy expert receive private signals about the state of the world, the institution sends a public signal, and a government must decide whether to implement a policy reform or delegate policy to an expert who has superior information but may not share its preferences. We conclude that persuasion is possible, but only when there is an expected preference difference between the IO and domestic experts, and only when the international institution holds the more moderate policy position. In such cases, persuasion takes place on two levels. First, the international institution can send a credible signal about a crisis and prompt the government to take action in response, and second, it can direct the government's attention to domestic experts and make their expertise policy relevant.

The intuition for our results is straightforward: when the policy the expert chooses is exactly what the IO would like it to choose, the IO has an incentive to misrepresent its private information in order to convince the government to delegate authority. The IO can only credibly communicate its private information if it shares a government's concern about the bias of its domestic experts. This suggests that, far from being an obstacle to international cooperation, polarized domestic politics may be a necessary condition for international organizations to exert effective influence. In fact, when domestic experts move into the political mainstream, international organizations may no longer be able to credibly signal the need for reform.

When international organizations are able to influence domestic politics in our model,

they do so by promoting the careers of elite experts with preferences that differ from those of the government. For example, environmental IOs promote the careers of environmental scientists who prefer activist policies to reverse pollution and retard global climate change, and economic IOs promote the careers of economists who prefer reformist policies. International organizations can be said to influence the “preferences” of the government in the sense that they promote the careers of like-minded officials, which changes the baseline for future policy interactions. In this way, the interaction between IOs and domestic bureaucratic politics can lead to a gradual convergence of policies, perspectives and expectations that builds a community of interests, even if some of the participating countries are authoritarian.

Finally, the results have significant implications for the design and reform of international organizations. International organizations facilitate cooperation by persuading member governments to take actions that are in their own interests, and this is only possible if they can establish trust. Trust is inherently problematic, because international organizations aggregate their members’ interests, which are unlikely to be aligned with those of a particular member during a crisis. Our results indicate that persuasion and trust are possible only when the international organization takes a position that is moderate with respect to the policy alternatives provided by domestic politics. International organizations can exert effective persuasion only within a relatively narrow range, and effectiveness demands attention to domestic politics.

Appendix

Proposition 1. *If there are only unbiased domestic experts, no truth-telling equilibrium exists in which the government conditions its decision whether to delegate on the institution's message.*

Proof. Suppose I truthfully reports its signal, i.e. $m_I = s_I \in \{0, 1\}$. We show that for some parameter range of k this cannot be an equilibrium.

(i) Suppose $m_I = 0$ and it is public knowledge. Then the players' updated belief that $\theta = 1$ is $\mu_0 = Pr(\theta = 1 | m_I = 0) = 1 - \alpha$.⁶ If G chooses policy itself, then the policy is the solution to the following maximization problem:

$$\max_{x \geq 0} EU_G(x | m_I = 0) = (1 - \mu_0)(-x) + \mu_0(x - 1) - kx$$

The optimal solution is

$$x_G = \begin{cases} 0 & \text{if } k \geq 2\mu_0 - 1, \\ 1 & \text{if } k < 2\mu_0 - 1. \end{cases} \quad (3)$$

Because $\mu_0 < 1/2$ and $k > 0$, G will choose $x = 0$. Now consider G 's decision whether to delegate the policy choice to E . Given $s_I = 0$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha\beta + (1 - \alpha)(1 - \beta)$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha + \beta - 2\alpha\beta$. For the ease of exposition, let $\mu_{i,j}$ denote E 's belief about the state of the world being 1 given $s_E = i$ and $m_I = j$, and $x_{i,j}$ denote E 's policy choice given $\mu_{i,j}$. If E receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,0} = Pr(\theta = 1 | s_E = 0, m_I = 0) = \frac{(1-\alpha)(1-\beta)}{(1-\alpha-\beta+2\alpha\beta)} < 1/2$, and it will recommend $x_{0,0} = 0$. If E receives $s_E = 1$, her updated belief about the state of the world is $\mu_{1,0} = Pr(\theta = 1 | s_E = 1, m_I = 0) = \frac{(1-\alpha)\beta}{\alpha+\beta-2\alpha\beta} > 1/2$, in which case she will choose policy $x_{1,0} = 1$. Essentially, because the quality of the expert's signal is better than the institution's, the expert will choose a policy that is consistent with its own signal. As a

⁶In all proofs, the subscript of μ will refer to the signal reported by I and E .

result, by delegating the policy to E the expected utility for G is:

$$\begin{aligned}
EU_G(E, m_I = 0) &= (1 - \alpha - \beta + 2\alpha\beta)[-(1 - \mu_{0,0})x_{0,0} + \mu_{0,0}(x_{0,0} - 1) - kx_{0,0}] \\
&\quad + (\alpha + \beta - 2\alpha\beta)[-(1 - \mu_{1,0})x_{1,0} + \mu_{1,0}(x_{1,0} - 1) - kx_{1,0}] \\
&= -(1 - \alpha - \beta + \alpha\beta) - \alpha(1 - \beta) - (\alpha + \beta - 2\alpha\beta)k \\
&= \beta - 1 - (\alpha + \beta - 2\alpha\beta)k
\end{aligned} \tag{4}$$

G will compare this expected utility with that from choosing a policy itself after I 's message, $EU_G(\neg E, m_I = 0) = \alpha - 1$. G will choose the action that gives it a higher expected utility of the two. We find that if $k < \frac{\beta - \alpha}{\alpha + \beta - 2\alpha\beta}$, then G will delegate. In other words, if the cost of implementing the policy is large, then G will avoid delegating because E may choose $x = 1$. If the cost is small, then G is willing to delegate for a better policy outcome which will overweigh the cost of implementing the policy.

(ii) Suppose $m_I = 1$. Then G knows $\theta = 1$ with $\mu_1 = Pr(\theta = 1 | m_I = 1) = \alpha$. If G chooses policy itself, then the policy is the solution to the following maximization problem:

$$\max_{x \geq 0} EU_G(x | m_I = 1) = (1 - \mu_1)(-x) + \mu_1(x - 1) - kx$$

The optimal solution is

$$x_G = \begin{cases} 0 & \text{if } k \geq 2\mu_1 - 1, \\ 1 & \text{if } k < 2\mu_1 - 1. \end{cases} \tag{5}$$

Because $\mu_1 > 1/2$, G can either choose $x = 0$ or $x = 1$ depending on the size of k . For $k \geq 2\alpha - 1$, G will choose $x = 0$.

Now consider G 's decision whether to delegate the policy choice to E . Given $s_I = 1$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha + \beta - 2\alpha\beta$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha\beta + (1 - \alpha)(1 - \beta)$. If E receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,1} = Pr(\theta = 1 | s_E = 0, m_I = 1) = \frac{\alpha(1 - \beta)}{\alpha + \beta - 2\alpha\beta} < 1/2$, and it will recommend $x_{0,1} = 0$. If E receives $s_E = 1$, her updated belief about the state of the world

is $\mu_{1,1} = Pr(\theta = 1 | s_E = 1, m_I = 1) = \frac{\alpha\beta}{1-\alpha-\beta+2\alpha\beta} > 1/2$, in which case she will choose policy $x_{1,1} = 1$. As a result, by delegating the policy to E the expected utility for G is:

$$\begin{aligned}
EU_G(E, m_I = 1) &= (\alpha + \beta - 2\alpha\beta)[-(1 - \mu_{0,1})x_{0,1} + \mu_{0,1}(x_{0,1} - 1) - kx_{0,1}] \\
&\quad (1 - \alpha - \beta + 2\alpha\beta)[-(1 - \mu_{1,1})x_{1,1} + \mu_{1,1}(x_{1,1} - 1) - kx_{1,1}] \\
&= -\alpha(1 - \beta) - (1 - \alpha - \beta + \alpha\beta) - (1 - \alpha - \beta + 2\alpha\beta)k \\
&= -1 + \beta - (1 - \alpha - \beta + 2\alpha\beta)k
\end{aligned} \tag{6}$$

G will compare this expected utility with that from choosing a policy itself after I 's message. Suppose $k \geq 2\alpha - 1$, then $EU_G(\neg E, m_I = 1) = -\alpha$. G will choose the action that gives it a higher expected utility of the two. We find that if $k \geq \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta} > 2\alpha - 1$, then G will choose its own policy; if $2\alpha - 1 \leq k < \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}$, then G will delegate. In other words, if the cost of implementing the policy is large, then G will avoid delegating because E may choose $x = 1$. Suppose $k < 2\alpha - 1$, then $EU_G(\neg E, m_I = 1) = \alpha - 1 - k$. We find that G will choose its own policy if $k \leq \frac{\alpha-\beta}{\alpha+\beta-2\alpha\beta}$. The condition cannot be true because $\beta > \alpha$, therefore, in this case G will always delegate. In sum, if $k < \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}$, then G delegates. Compared with the case $m_I = 0$, for a larger range of k , G will delegate.⁷

Given G 's strategy, does I have an incentive to deviate from the truth-telling strategy?

Suppose $s_I = 0$. If I reports truthfully so that $m_I = 0$, then G will choose its own policy if $k \geq \frac{\beta-\alpha}{\alpha+\beta-\alpha\beta}$. If I deviates from truth-telling and sends $m_I = 1$ to G , then G will choose its own policy if $k \geq \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}$. Because the quality of E 's information is higher than that of the institution's, I is better off when G delegates regardless of the signal that it received ($\beta - 1 > \alpha - 1 > -\alpha$). Let $k_1 = \frac{\beta-\alpha}{\alpha+\beta-\alpha\beta}$ and $k_2 = \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}$. Now we consider three scenarios of the cost:

1. $k \leq k_1$, then G will delegate receiving either message from I , so I has no incentive to lie;

⁷ $\max\{\frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}, \frac{\beta-\alpha}{\alpha+\beta-2\alpha\beta}\} = \frac{\alpha+\beta-1}{1-\alpha-\beta+2\alpha\beta}$.

2. $k_1 < k < k_2$, then G will not delegate when $m_I = 0$, but will delegate when $m_I = 1$, so I has an incentive to lie;
3. $k \geq k_2$, then G will not delegate in either case. So there is no incentive for I to lie.

As a result, I has an incentive to lie if $k_1 \leq k < k_2$.

Suppose $s_I = 1$. Again we have three cases to consider and only when $k_1 \leq k < k_2$, G will delegate contingent on I 's message. But in this case I has no incentive to lie because G will delegate if I reports truthfully, and will not delegate if I reports $m_I = 0$. As a result, I will always tell the truth when $s_I = 1$.

In sum, when $\frac{\beta - \alpha}{\alpha + \beta - \alpha\beta} \leq k < \frac{\alpha + \beta - 1}{1 - \alpha - \beta + 2\alpha\beta}$, there is no truth telling equilibrium; for other values of k , a truth-telling equilibrium exists. \square

Proposition 2. *If there exist biased experts who always prefer taking a high action, and the probability of encountering such experts is sufficiently high, then for some parameter range, there exists a truth-telling equilibrium in which the institution truthfully reveals its information and the government decides whether to delegate the policy on the information.*

Proof. Suppose with π the expert is biased toward action 1. Because the expert gets to choose the policy if G delegates, both types of expert will choose the policy according to her true preference. This means that the biased type always chooses $x_E = 1$. On the other hand, the unbiased type will have an incentive to recommend a policy that is consistent with her updated belief about the state of the world.

Consider the equilibrium in which I truthfully reports its signal, i.e. $m_I = s_I \in \{0, 1\}$.

(i) Suppose $m_I = 0$ and it is public knowledge. Then the players' updated belief that $\theta = 1$ is $\mu_0 = Pr(\theta = 1 | m_I = 0) = 1 - \alpha$.⁸ If G chooses policy by itself, then the policy is

⁸In all proofs, the subscript of μ will refer to the signal reported by I and E .

the solution to the following maximization problem:

$$\max_{x \geq 0} EU_G(x|m_I = 0) = (1 - \mu_0)(-x) + \mu_0(x - 1) - kx$$

The optimal solution is

$$x_G = \begin{cases} 0 & \text{if } k \geq 2\mu_0 - 1, \\ 1 & \text{if } k < 2\mu_0 - 1. \end{cases} \quad (7)$$

Because $\mu_0 < 1/2$ and $k > 0$, G will choose $x = 0$. Now consider G 's decision whether to delegate the policy choice to E . G knows that with $1 - \pi$ the expert is unbiased and will choose a policy that is consistent with her updated belief about the state of the world; with π the expert is biased and will recommend simply $x = 1$ regardless of the signal she receives. Consider the decision by a unbiased expert. Given $s_I = 0$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha\beta + (1 - \alpha)(1 - \beta)$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha + \beta - 2\alpha\beta$. If E receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,0} = Pr(\theta = 1|s_E = 0, m_I = 0) = \frac{(1-\alpha)(1-\beta)}{(1-\alpha-\beta+2\alpha\beta)} < 1/2$, and it will recommend $x_{0,0} = 0$. If E receives $s_E = 1$, her updated belief about the state of the world is $\mu_{1,0} = Pr(\theta = 1|s_E = 1, m_I = 0) = \frac{(1-\alpha)\beta}{\alpha+\beta-2\alpha\beta} > 1/2$, in which case she will choose policy $x_{1,0} = 1$. Again, given that the quality of the expert's signal is higher than the institution's, the expert will choose a policy that is consistent with its own signal. As a result, by delegating the policy to E the expected utility for G is:

$$\begin{aligned} EU_G(E, m_I = 0) &= (1 - \alpha - \beta + 2\alpha\beta)\{(1 - \mu_{0,0})[-(1 - \pi)(x_{0,0} + kx_{0,0}) - \pi(1 + k)] \\ &\quad + \mu_{0,0}[(1 - \pi)(x_{0,0} - 1 - kx_{0,0}) - \pi k]\} \\ &\quad + (\alpha + \beta - 2\alpha\beta)\{(1 - \mu_{1,0})[-(1 - \pi)(x_{1,0} + kx_{1,0}) - \pi(1 + k)] \\ &\quad + \mu_{1,0}[(1 - \pi)(x_{1,0} - 1 - kx_{1,0}) - \pi k]\} \\ &= -\alpha\beta\pi(1 + k) - (1 - \alpha)(1 - \beta)(1 - \pi + \pi k) - \alpha(1 - \beta)(1 + k) - \beta(1 - \alpha)k \end{aligned} \quad (8)$$

G will compare this expected utility with that from choosing a policy itself after I 's message, $EU_G(\neg E, m_I = 0) = \alpha - 1$. Because having the biased expert is even more likely to lead

to policy $x = 1$, the threshold value of k that will give G an incentive to delegate is going to be smaller than the case where there is no biased expert. We find the threshold to be $k'_1 = \frac{\beta - \alpha - (\alpha + \beta - 1)\pi}{\alpha\beta\pi + (1 - \alpha)(1 - \beta)\pi + \alpha(1 - \beta) + \beta(1 - \alpha)}$, and $k_1 > k'_1$. Moreover, if $\pi < \frac{\beta - \alpha}{\alpha + \beta - 1}$, then $k'_1 > 0$.⁹ That is, if π is small enough, then G will choose a policy by itself if $k \geq k'_1$. This result is similar to the case of unbiased expert. If $\pi > \frac{\beta - \alpha}{\alpha + \beta - 1}$, then $k'_1 < 0$, which means for any positive k , G will choose its own policy.

(ii) Suppose $m_I = 1$. Then G knows $\theta = 1$ with $\mu_1 = Pr(\theta = 1 | m_I = 1) = \alpha$. If G chooses policy by itself, then the policy is the solution to the following maximization problem:

$$\max_{x \geq 0} EU_G(x | m_I = 1) = (1 - \mu_1)(-x) + \mu_1(x - 1) - kx$$

The optimal solution is

$$x_G = \begin{cases} 0 & \text{if } k \geq 2\mu_1 - 1, \\ 1 & \text{if } k < 2\mu_1 - 1. \end{cases}$$

Because $\mu_1 > 1/2$, G can either choose $x = 0$ or $x = 1$ depending on the size of k . For $k \geq 2\alpha - 1$, G will choose $x = 0$.

Now consider G 's decision whether to delegate the policy choice to E . Given $s_I = 1$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha + \beta - 2\alpha\beta$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha\beta + (1 - \alpha)(1 - \beta)$. Consider the decision by an unbiased expert. If E receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,1} = Pr(\theta = 1 | s_E = 0, m_I = 1) = \frac{\alpha(1 - \beta)}{\alpha + \beta - 2\alpha\beta} < 1/2$, and it will recommend $x_{0,1} = 0$. If E receives $s_E = 1$, her updated belief about the state of the world is $\mu_{1,1} = Pr(\theta = 1 | s_E = 1, m_I = 1) = \frac{\alpha\beta}{1 - \alpha - \beta + 2\alpha\beta} > 1/2$, in which case she will choose policy $x_{1,1} = 1$. As a result, by delegating the policy to E the expected utility for G is:

$$EU_G(E, m_I = 1) = (\alpha + \beta - 2\alpha\beta)\{(1 - \mu_{0,1})[-(1 - \pi)(x_{0,1} + kx_{0,1}) - \pi(1 + k)]$$

⁹It can be shown that $0 < \frac{\beta - \alpha}{\alpha + \beta - 2\alpha\beta} < 1$.

$$\begin{aligned}
& +\mu_{0,1}[(1-\pi)(x_{0,1}-1-kx_{0,1})-\pi k]\} \\
& +(1-\alpha-\beta+2\alpha\beta)\{(1-\mu_{1,1})[-(1-\pi)(x_{1,1}+kx_{1,1})-\pi(1+k)] \\
& \quad +\mu_{1,1}[(1-\pi)(x_{1,1}-1-kx_{1,1})-\pi k]\} \\
& = -\beta(1-\alpha)\pi(1+k) - \alpha(1-\beta)(1-\pi+\pi k) - (1-\alpha-\beta+\alpha\beta)(1+k) - \alpha\beta k \tag{9}
\end{aligned}$$

G will compare this expected utility with that from choosing a policy by itself after I 's message.

Suppose $k \geq 2\alpha - 1$, then G will choose $x = 0$ if it chooses the policy by itself, and $EU_G(\neg E, m_I = 1) = -\alpha$. G will choose the action that gives it a higher expected utility of the two. We find the threshold to be $k'_2 = \frac{\alpha+\beta-1+(\alpha-\beta)\pi}{(\alpha+\beta-2\alpha\beta)\pi+(1-\alpha-\beta+2\alpha\beta)}$, and $k_2 > k'_2$. Moreover, if $\pi < \frac{\alpha+\beta-1}{\beta-\alpha}$, then $k'_2 > 0$. It can be shown that $\frac{\alpha+\beta-1}{\beta-\alpha} > 1$, so $k'_2 > 0$, and also $k_2 > k'_2$ as expected. It then follows if $k \geq k'_2$, then G will choose the policy by itself, and if $2\alpha - 1 < k < k'_2$, then it will delegate to the expert.¹⁰

Now suppose $k < 2\alpha - 1$, then G will choose $x = 1$ if it chooses the policy by itself, and $EU_G(\neg E, m_I = 1) = \alpha - 1 - k$. We find that G will choose its own policy if $k \leq \frac{\alpha-\beta}{\alpha+\beta-2\alpha\beta}$. The condition cannot be true because $\beta > \alpha$, therefore, in this case G will always delegate. In sum, if $k < k'_2$, then G delegates. Compared with the case $m_I = 0$, for a larger range of k , G will delegate.¹¹

Given G 's strategy, does I have an incentive to deviate from the truth-telling strategy? We first consider the case that $\pi \leq \frac{\beta-\alpha}{\alpha+\beta-1}$, so that $0 < k'_1 < k'_2$.

Suppose $s_I = 0$. If I reports truthfully so that $m_I = 0$, then G will take that information and choose its own policy if $k \geq k'_1$. If I deviates from truth-telling and sends $m_I = 1$ to G , then G will choose its own policy if $k \geq k'_2$. With the existence of the biased type, I is not always better off from policy delegation.

¹⁰It can be shown $2\alpha - 1 < k'_2$ for all $\pi > 0$.

¹¹ $\max\{k'_1, k'_2\} = k'_2$.

1. $k \leq k'_1$, then G will delegate receiving either message from I , so I has no incentive to lie;
2. $k'_1 < k < k'_2$, then G will not delegate when $m_I = 0$, but will delegate when $m_I = 1$. This is a case where I may have an incentive to lie. However, if π is sufficiently large, then I has no incentive to mislead G to delegate because of a high risk of getting a biased expert. To see that this is the case, consider I 's utility from being truthful and lying. If I is truthful, then G will not delegate and choose policy 0. Then I 's expected utility is $\alpha - 1$. If I lies, then G will delegate to the expert. Given $s_I = 0$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha\beta + (1 - \alpha)(1 - \beta)$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha + \beta - 2\alpha\beta$. If an unbiased expert receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,0} = Pr(\theta = 1|s_E = 0, m_I = 0) = \frac{(1-\alpha)(1-\beta)}{(1-\alpha-\beta+2\alpha\beta)} < 1/2$, and she will choose policy $x_{0,0} = 0$; if she receives $s_E = 1$, her updated belief about the state of the world is $\mu_{1,0} = Pr(\theta = 1|s_E = 1, m_I = 0) = \frac{(1-\alpha)\beta}{\alpha+\beta-2\alpha\beta} > 1/2$, and she will choose policy $x_{1,0} = 1$. A biased expert will always choose $x = 1$. The expected utility for I in this case is:

$$\begin{aligned}
EU_I(m_I = 1|s_I = 0) &= (1 - \alpha - \beta + 2\alpha\beta)\{(1 - \mu_{0,0})[-(1 - \pi)x_{0,0} - \pi] + \mu_{0,0}(1 - \pi)(x_{0,0} - 1)\} \\
&\quad + (\alpha + \beta - 2\alpha\beta)\{(1 - \mu_{1,0})[-(1 - \pi)x_{1,0} - \pi] + \mu_{1,0}(1 - \pi)(x_{1,0} - 1)\} \\
&= (1 - \alpha - \beta + 2\alpha\beta)\{-(1 - \mu_{0,0})\pi - \mu_{0,0}(1 - \pi)\} + (\alpha + \beta - 2\alpha\beta)\{-(1 - \mu_{1,0})\} \\
&= -\alpha\beta\pi - (1 - \alpha)(1 - \beta)(1 - \pi) - \alpha(1 - \beta) \\
&= \beta - 1 + \pi(1 - \alpha - \beta)
\end{aligned}$$

We find that if $\pi \geq \frac{\beta - \alpha}{\alpha + \beta - 1}$, I is better off being truthful and let G choose the policy. This contradicts the assumption $\pi \leq \frac{\beta - \alpha}{\alpha + \beta - 1}$ for this case. Therefore, I will always have an incentive to lie if the probability of drawing a biased expert is not too high.¹²

3. $k \geq k'_2$, then G will not delegate in either case. So there is no incentive for I to lie.

¹²When $\pi = \frac{\beta - \alpha}{\alpha + \beta - 1}$, I is indifferent and we assume that I will be truthful.

As a result, I has an incentive to lie when $k'_1 < k < k'_2$, but will not do so if π is sufficiently high.

Suppose $s_I = 1$. From the above analysis we know that I only has an incentive to lie if $k'_1 < k < k'_2$. If I lies, then G will not delegate and choose policy 0. Then I 's expected utility is $-\alpha$. If I reports truthfully, then G will delegate. Given $s_I = 1$, the expert will receive $s_E = 0$ with $Pr(s_E = 0) = \alpha + \beta - 2\alpha\beta$ and $s_E = 1$ with $Pr(s_E = 1) = \alpha\beta + (1 - \alpha)(1 - \beta)$. If an unbiased expert receives $s_E = 0$, her updated belief about the state of the world being 1 is $\mu_{0,1} = Pr(\theta = 1 | s_E = 0, m_I = 1) = \frac{\alpha(1-\beta)}{\alpha+\beta-2\alpha\beta} < 1/2$, and it will recommend $x_{0,1} = 0$; if she receives $s_E = 1$, her updated belief about the state of the world is $\mu_{1,1} = Pr(\theta = 1 | s_E = 1, m_I = 1) = \frac{\alpha\beta}{1-\alpha-\beta+2\alpha\beta} > 1/2$, and she will choose policy $x_{1,1} = 1$. As a result, by delegating the policy to E the expected utility for I is:

$$\begin{aligned} EU_I(E, m_I = 1) &= (\alpha + \beta - 2\alpha\beta)\{(1 - \mu_{0,1})[-(1 - \pi)(x_{0,1}) - \pi] + \mu_{0,1}(1 - \pi)(x_{0,1} - 1)\} \\ &\quad + (1 - \alpha - \beta + 2\alpha\beta)\{(1 - \mu_{1,1})[-(1 - \pi)x_{1,1} - \pi] + \mu_{1,1}(1 - \pi)(x_{1,1} - 1)\} \\ &= -\beta(1 - \alpha)\pi - \alpha(1 - \beta)(1 - \pi) - (1 - \alpha - \beta + \alpha\beta) \\ &= \beta - 1 + \pi(\alpha - \beta) \end{aligned}$$

For all $1 \geq \pi \geq 0$, I 's expected utility from delegating is better than not delegating, so I has no incentive to lie in this case.

Now consider the case $\pi > \frac{\beta - \alpha}{\alpha + \beta - 1}$, so that $k'_1 < 0$. Then there are only two scenarios to discuss.

1. $0 < k < k'_2$. G will not delegate if $m_I = 0$, but delegate if $m_I = 1$. Suppose $s_I = 0$. From the previous analysis we find that if $\pi > \frac{\beta - \alpha}{\alpha + \beta - 1}$, then I is better off reporting $m_I = 0$ and let G choose the policy. Therefore, I will be truthful in this case. Suppose $s_I = 1$. From the previous analysis we know that for all $1 \geq \pi \geq 0$, I receives a higher payoff from policy delegation, so I will be truthful.

2. $k \geq k'_2$, then G will not delegate for either message, so I has no incentive to lie.

In sum, with the existence of the biased type, I has an incentive to lie when receiving signal $s_I = 0$, but if $\pi > \frac{\beta - \alpha}{\alpha + \beta - 1}$, i.e., the probability of encountering a biased expert is sufficiently high, then there will be a truth-telling equilibrium for all k . \square

Proposition 3. *If there exist biased experts who always prefer taking a low action, then there does not exist a truth-telling equilibrium in which the government conditions its decision whether to delegate on the institution's message.*

The proof is similar to that of Proposition 2 and is omitted due to space consideration.

References

- [1] Aghion, Philippe and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105(1):1-29.
- [2] Abdelal, Rawi. 2007. *Capital Rules: The Construction of Global Finance*. Cambridge, Mass.: Harvard University Press.
- [3] Aleksashenko, Sergei. 1999. *Bitva za rubl'*. Moscow: Alma Mater.
- [4] Anders Aslund. 1995. *How Russia Became a Market Economy* Washington, D.C.: Brookings Institution.
- [5] Bendor, Jonathan, and Adam Meirowitz. 2004. "Spatial Models of Delegation." *American Political Science Review* 98(2):293-310.
- [6] Bendor, Jonathan, and Adam Meirowitz. 2004. "Spatial Models of Delegation." *American Political Science Review* 98(2):293-310.
- [7] Boughton, James M. 2001. *Silent Revolution: The International Monetary Fund, 1979-1989*. Washington, D.C.: International Monetary Fund.
- [8] Calvert, Randall. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47:530-555.
- [9] Chapman, Terrence. 2009. "Audience Beliefs and International Organization Legitimacy." *International Organization* 63(4): 733-764.
- [10] Checkel, Jeffrey T. 1997. *Ideas and international political change : Soviet/Russian behavior and the end of the Cold War*. New Haven : Yale University Press.
- [11] Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):
- [12] Cukierman, Alex and Mariano Tommasi. 1998. "When Does it Take a Nixon to Go to China?" *The American Economic Review* 88(1):180-197.
- [13] Dai, Xinyuan. 2005. "Comply? The Domestic Constituency Mechanism." *International Organization* 59 (Spring):363-398.
- [14] Dessein, Wouter. 2002. "Authority and Communication in Organizations." *Review of Economic Studies* 69:811-838.
- [15] Drazen, Allan. 2002. Conditionality and Ownership in IMF Lending: A Political Economy Approach. CEPR Discussion Paper No. 3562.
- [16] Drazen, Allen and Paul Masson. 1994. "Credibility of Policies Versus Credibility of Policy-Makers." *The Quarterly Journal of Economics* 109(3):735-754.
- [17] Dreher, Axel, and Nathan M. Jensen. 2007. "Independent Actor or Agent? An Empirical Analysis of the Impact of US Interests on IMF Conditions." *Journal of Law and Economics* 50.

- [18] Dreher, Axel, and Roland Vaubel. 2004. The Causes and Consequences of IMF Conditionality. *Emerging Markets Finance and Trade* 40 (3): 26-54.
- [19] Fang, Songying. 2008. "The Informational Role of International Institutions and Domestic Politics." *American Journal of Political Science* 52(2):304-321.
- [20] Fearon, James D. 1995. "Rationalist Explanations for War." *International Organization* 49(Summer): 379-414.
- [21] Gailmard, Sean. 2009. "Discretion Rather than Rules: Choice of Instruments to Control Bureaucratic Policy Making." *Political Analysis* 17: 25-44.
- [22] Gilligan, Thomas, and Keith Krehbiel. 1987. "Collective Decision-making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, and Organization* 3:287-335.
- [23] Goldstein, Judith and Robert O. Keohane. 1993. *Ideas and Foreign Policy: Beliefs, Institutions, and Political Change*. Ithaca: Cornell University Press.
- [24] Gould, Erica R. 2003. Money Talks: Supplementary Financiers and International Monetary Fund Conditionality. *International Organization* 57 (Summer): 551-86.
- [25] Gould, Erica R. 2006. *Money Talks: The International Monetary Fund, Conditionality, and Supplementary Financiers*. Stanford: Stanford University Press.
- [26] Helleiner, Eric. *States and the Reemergence of Global Finance: From Bretton Woods to the 1990s*. (Ithaca: Cornell Univ. Press, 1994).
- [27] Hills, Carla A., Peter G. Peterson and Morris Goldstein 1999. *Safeguarding Prosperity in a Global Financial System: The Future International Financial Architecture*. Washington, DC: Council on Foreign Relations and Institute for International Economics.
- [28] International Monetary Fund. 2005. Review of the 2002 Conditionality Guidelines. Washington D.C. <http://www.imf.org/external/np/pp/eng/2005/030305.pdf>.
- [29] Independent Evaluation Office, International Monetary Fund. 2003. *The IMF and Recent Capital Account Crises: Indonesia, Korea, Brazil*. Washington, D.C.: International Monetary Fund.
- [30] Independent Evaluation Office, International Monetary Fund. 2004. *The IMF and Argentina, 1991-2001*. Washington, D.C.: International Monetary Fund.
- [31] Keohane, Robert O. 1984. *After Hegemony*. Princeton, NJ: Princeton University Press.
- [32] Khan, Mohsin S. and Sharma, Sunil. 2001. IMF Conditionality and Country Ownership of Programs. IMF Working Paper No. 01/142.
- [33] Kydd, Andrew H. 2005. *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.
- [34] Kydd, Andrew H. 2003. "Which Side Are You On? Bias, Credibility and Mediation." *American Journal of Political Science* 47 (4): 597-611.

- [35] Mansfield, Edward D., Helen V. Milner, and B. Peter Rosendorff. 2002. "Democracies Cooperate More: Electoral Control and International Trade Agreements." *International Organization* 56(3):477-513.
- [36] Milner, Helen V. 1997. *Interests, Institutions, and Information: Domestic Politics and International Relations*. Princeton: Princeton Univ. Press.
- [37] Milner, Helen V. 1997. *Interests, Institutions, and Information Trust*. Princeton, NJ: Princeton University Press.
- [38] Mussa, Michael. 2002. *Argentina and the Fund: From Triumph to Tragedy*. Washington, D.C.: Institute for International Economics (July).
- [39] Pauly, Louis W. 1997. *Who Elected the Bankers? Surveillance and Control in the World Economy*. Ithaca: Cornell University Press.
- [40] Sartori, Anne E. 2005. *Deterrence by diplomacy*. Princeton, N.J. : Princeton University Press,
- [41] Woods, Ngaire. 2006. *The Globalizers : The IMF, the World Bank, and their Borrowers*. Ithaca: Cornell University Press.