The Problem of Insincere Compliance in International Relations: Norms, Policy Diffusion, and International Expectations[1]

Susan D. Hyde

Associate Professor of Political Science and International Affairs

Yale University

Susan.hyde@yale.edu

March 20, 2015

Draft in progress

Please do not cite or circulate.

[1] The paper draws heavily on some prior work (Hyde 2011a, 2011b) in order to extend a previously published theory to other international norms.

For many sovereign states, the post-WWII period of globalization has been accompanied by increasing homogenization of policies and practices across dozens of issue areas in political, economic, and military affairs. For example, by 2010, the vast majority of sovereign states had adopted an independent central bank; signed onto the international regulation of their nuclear program; committed to numerous bilateral investment treaties; and invited sovereign bond ratings by privately run international agencies. Many states had also adopted a range of institutions associated with democratic or democratizing states, including national elections, formal electoral competition, an ostensibly free media, universal adult suffrage, and an independent electoral commission. Although many of these behaviors and policies have been widely adopted by states, with some of them becoming nearly universal, in most issue areas a subset of states adopt these new policies disingenuously, with no intention of making the deeper changes that one might expect should accompany genuine reform.

This problem of insincere compliance is not yet widely recognized as a more general phenomenon in international relations, even though such issues have been highlighted across many issue areas. Many leaders who adopt national elections have no intention of stepping down in the event of an election loss, regimes that adopt gender quota for women may not be genuinely committed to improvements in women's rights (Bush 2011), governments that move towards formal independence of central banks may still exert political pressure on them (Franzese 1999), and governments make a variety of policy changes aimed at attracting foreign direct investment and reducing corruption with no real intention of reform.

Motivated by these global trends, this paper focuses on two related questions: Why do state-level policies and practices diffuse and become internationally expected behaviors, or international norms? And why do these policies and practices often fail to have their intended effects?

I build on existing research by linking explanations of norm diffusion with those of international policy diffusion, and provide a potential resolution to initially contradictory findings about the effectiveness of widely adopted policies and practices. This theory has the potential to apply to diverse issue areas including democratic institutions, central bank independence, bilateral investment treaties, weapons inspection, and treaties on human rights and nuclear nonproliferation. I explain the diffusion of policies and practices—and their sometimes heterogeneous effects –by emphasizing the role of shared expectations among international

actors about the observable behaviors of other states, and considering how these shared expectations influence the reasons why states adopt diffusing policies.

The preliminary theory presented in this paper is based on the assumption that states and other international actors allocate benefits to other states, and that they do so partly based on a desire to reward specific desirable characteristics, which may include things like stability, democracy, loyalty, transparency, a business-friendly investment climate, respect for human rights, likelihood of repaying sovereign debts, and many others. I also assume that states in the international system vary in whether they possess desirable characteristics. International actors value some characteristics of states more than others, and it can be difficult for international actors (including other sovereign states, multi-national corporations, foreign investors, and international organizations) to evaluate which states are actually "good" types. These characteristics may be promoted by international and domestic actors, and may be rewarded implicitly or explicitly. The characteristics valued by international actors change over time, and there is not necessarily a formal coordination mechanism among benefit-giving international actors when determining which characteristics are valued. Rather, from the perspective of a state seeking international benefits, the expected international value of any given policy or practice is often determined through a process of trial and error and experience over time.

In general, I suggest that benefit-seeking states with desirable characteristics are motivated to find credible signals of their type. These signals may be specific behaviors or policies, like holding national elections and allowing electoral competition, or allowing international weapons inspection to take place. For a behavioral signal to be credible to external (international) audiences, it must be more costly for a state to adopt if the state does not actually possess the internationally valued characteristic. If a prospective signal is successful in communicating a state's underlying characteristic to international audiences, it is rewarded by those international actors who value the underlying characteristic, such as democratization or compliance with weapons treaties. Note that a signal need not lead to perfect separation of types in order to be informative.

If a signal is successful at communicating a state's type, it is likely to be mimicked by other benefit-seeking actors, a dynamic which causes the new behavior to diffuse to other states that are seeking international benefits. States that actually possess the desirable characteristic have the easiest time sending the signal. The behavior becomes an international norm when

benefit-giving actors develop the belief that all 'good' governments (or true types) send the signal. If international actors believe that all true types adopt a given signal, even states that must fake the signal may be motivated to adopt it. Thus, under specified conditions, the policy or behavior that is used as the signal becomes expected behavior for all true types, may spread to bad types, and can become nearly universal even in the absence of explicit advocacy or pressure on states to adopt the new behavior.

This article first defines international norms, summarizes dominant explanations for international norm formation and policy diffusion, introduces the theory in general terms, and then discusses it in the context of several cases. It is intended to be primarily theoretical, with the issue areas discussed serving as illustrations of the argument rather than a formal test of hypotheses.

Defining and Clarifying International Norms

Although this article discusses international norms, it is not intended to evaluate or test any of the "isms" in international relations. Although international norms are, to date, primarily studied by constructivists (or by rationalists who tend to avoid using the term "international norm" even when referring to shared expectations that meet the definition of an international norm), this theory draws on both constructivist and rationalist theories of diffusion. As such, several clarifications are in order because the mention of international norms often provokes strong yet contradictory responses among scholars of international relations.

International norms are shared "standards of appropriate behavior for actors with a given identity" (Finnemore and Sikkink 1998). Note that this definition focuses on the shared expectations of behavior rather than the reasons that justify the shared expectation. Finnemore and Sikkink's widely used definition of international norm is similar to the concept of equilibrium beliefs in game theory. Some scholars question what the added value of using the term "international norm" might be when one could just say "shared expectations of behavior," or vice versa. On some level, I'm agnostic about the term used, but like many other scholars (Axelrod and Keohane 1985; Barnett 1998; Fearon and Wendt 2002; Finnemore and Sikkink 1998; Kelley 2008), I think that considering norms (or shared expectations) and rational action together provides a number of unrealized opportunities to better understand international political behavior.

There are several additional common misunderstandings amongst IR scholars that require clarification. First, whether norm compliance can be consistent with rational action is a common source of misconception about international norms. In fact, the debate about international norm formation is sometimes misperceived as a debate between rationalists and constructivists over whether norms matter. This perceived debate is outdated, at best, and by some accounts, it never took place. As many prominent scholars have emphasized, rational action is often an important part of norm compliance, and it is inaccurate to frame norm compliance and rationalism as alternatives (Fearon and Wendt 2002; Finnemore and Sikkink 1998; Katzenstein, Keohane, and Krasner 1998). Highlighting constructivist attention to rational action, Finnemore and Sikkink label the process of norm formation "strategic social construction," and argue that "rationality cannot be separated from any politically significant episode of normative influence or normative change, just as the normative context conditions any episode of rational choice." Peter Katzenstein, Robert Keohane, and Stephen Krasner argue that although constructivists and rationalists disagree on ontology, "on issues of epistemology and methodology, however, no great differences divide constructivists from rationalists." James Fearon and Alexander Wendt complain that rationalism and constructivism are often falsely pitted against one another, and some scholars mistakenly argue that "rationalists believe that people are always acting on material self-interest, and constructivists believe that people are always acting on the basis of norms and values." They go on to argue that this widely held misperception is due to misunderstanding of rationalism, not to any fundamental theoretical conflict between rationalism and constructivism.

A number of prominent scholars focused on international norms have highlighted that rational choice and constructivism complement each other more often than not, and that rational or strategic action is closely tied to norm initiation and norm compliance. For example, in relation to the norm of Arabism governing the international relations of Arab states, Michael Barnett argues that changes in this norm were generated through "social and strategic interactions" between rational and self-interested Arab states. In explaining state acceptance of the norm of territorial integrity, Mark Zacher argues that instrumental motivations played a large part in motivating states to accept the norm, in addition to democratic ideals promoted through international organizations (2003).

A second point of misunderstanding about international norms more generally is that many scholars perceive international norms as necessarily possessing a moral dimension. Note that a moral dimension is not part of the most widely cited modern definition of international norms summarized by Finnemore and Sikkink. Similarly, Goertz and Diehl (1992) highlight that the moral dimension is an important source of variation among international norms and a moral component is therefore not a necessary part of any international norm. To be sure, many international norms are clearly related to morality, such as international norms against slavery, torture, or child labor. These norms are important in influencing the behavior of states, and therefore have attracted the attention of scholars, most prominently in Keck and Sikkink's seminal book *Activists Without Borders* (Keck and Sikkink 1998). But the focus on morally-motivated activists has become so dominant in the political science literature on international norms that many consumers of this scholarship have developed the misperception that a moral component is a necessary component of an international norm.

This debate matters because there are a number of other behaviors that diffuse throughout the international system in part because of shared "expectations for actors with a given identity." Thus, international norms may be a fundamental part of explaining how diffusion takes place, as well as understanding the motivations of states which comply with the new norm.

In the next section I outline existing theories of policy diffusion and norm formation before outlining my theory, which represents an alternative causal mechanism of policy diffusion and norm formation. It is based on recent work explaining why international election observation became an international norm (Hyde 2011a, 2011b), but extends the theory in several ways.

Policy Diffusion and Norm Formation

How do existing theories explain global policy diffusion and international norm formation? Although these literatures have developed separately, their explanations for the diffusion of behaviors are similar. There are at least two paths to the creation of new international norms now identified in the existing literature. In the first theory, norms are generated because they encourage or reinforce mutually beneficial international cooperation. Individual states serve both as norm compliers and norm enforcers, and these cooperative norms tend to exist within broader sets of international institutions.

The second existing theory centers on norms initiated and spread by the work of norm entrepreneurs. These activist-centered norms are often intended to modify or prevent existing behaviors, such as the use of land mines, torture, child slavery, or nuclear weapons. The most prominent theory of norm development in international relations offers one possible explanation for the creation of norms under which compliance is costly. According to this theory, initiated by "constructivist" scholars of international relations, new and more controversial international norms are generated because coalitions of activists or powerful states pressure other actors to change their behavior. Activist pressure changes the decision calculus for leaders, and as a result state leaders are motivated to comply with the new norm. This activist-centered theory is most clearly articulated by Martha Finnemore and Kathryn Sikkink: In the first stage of norm initiation, "norm entrepreneurs attempt to convince a critical mass of states (norm leaders) to embrace new norms." Norm leaders then pressure other states to become norm followers, causing a "norm cascade" in which increasing numbers of states adopt the new norm. In the third stage of their theory, norms can become internalized and compliance with the norm becomes automatic (Finnemore and Sikkink 1998).

Although instrumental logics play a part in this theory and many related arguments— the work of activists may be intended to, for example, generate costs for actors who fail to comply with the new norm—norm entrepreneurs are central in initiating and spreading the new behavior. Without this pressure, the activist-centered theory implies that states or other international actors that are better off not complying with the potential norm would not be motivated to change their behavior. Therefore, unless advocates for a new norm are sufficiently powerful, influential, or persuasive, attempts to change state behavior and generate new international norms are unlikely to succeed.

For example, like Finnemore and Sikkink, Richard Price highlights the work of transnational activists in the global campaign to generate a norm against the use of anti-personnel landmines (Price 1998). Norm entrepreneurs and activists lobbied governments, generated international media attention against mine-producing states, mobilized domestic populations, and campaigned for the United Nations treaty banning the production and sale of land mines. Similarly, Nina Tannenwald argues that in addition to nuclear deterrence, a post-WWII international norm against the use of nuclear weapons explains the absence of nuclear weapon use since 1945 (Tannenwald 1999). This norm was created, she argues, in part because of the

work of a global and morally-motivated network of activists who campaigned against the use of nuclear weapons and raised moral objections against them. The work of activists, which may include networks of committed individuals, NGOS, or states, is intended to "mobilize popular opinion and political support both within their host country and abroad," and ultimately to motivate international actors to change their behavior (Nadelmann 1990, 482).

Finnemore and Sikkink's theory is so widely accepted as the dominant explanation for norm development in the literature that when encountering a new international norm, some scholars assume that it must have been generated through activist pressure. This is clearly a tautology that Finnemore and Sikkink did not intend, but highlights the dominant influence of their theory.

Advocacy-based theories of norm creation have been useful in explaining a number of now-prominent international norms governing state behavior, including the targeting of civilians in war, the production and use of land mines, trafficking in slaves, as well as those international norms proscribing individual behavior within states, such as international norms against child labor, killing endangered species, or discriminating against ethnic and religious minorities. They have also been useful in demonstrating that international norms "matter" and that states may comply with international norms even when it is costly for them to do so. Yet the activist theory of norm development does not—nor was it intended to—explain all international norms, nor all norms that are consequential. Finnemore and Sikkink argue that "norms do not appear out of thin air; they are actively built by agents having strong notions about appropriate or desirable behavior in their community." It is this feature of the Finnemore and Sikkink theory that most clearly distinguishes it from the argument offered here. By providing a theory of norm initiation and diffusion that does not require activism or imposition by powerful states, I am focusing explicitly on norms that are generated primarily through diffusely motivated strategic action, and that can be created even in the absence of activist pressure. By extending the logic of norm diffusion to these other issue areas in which activism is less likely to play a prominent role, I will also link norm diffusion more explicitly with policy diffusion.

Recent theories of policy diffusion also focus on instrumental motivations in explaining the spread of behaviors among states. For example, Beth Simmons and Zachary Elkins argue that the diffusion of neoliberal economic policies, including capital account liberalization, exchange rate policy unification, and current account liberalization, have taken place in part due to

international factors that influence information and the available set of policy choices (Simmons and Elkins 2004). They argue that the incentives for a given state to adopt a particular policy are influenced by the foreign policy choices of other states, and the information used by governments to make policy choices is also altered by policy choices in other states. Similarly, Simmons, Frank Dobbin, and Geoffrey Garrett theorize that policies diffuse between states via four processes: coercion, competition, learning, and emulation (Simmons, Dobbin, and Garrett 2006).

Competition, learning and emulation are all arguably elements of the signaling model of norm formation outlined in this paper, although the theory can be considered a more specific version of a diffusion model, and one that outlines a clear causal mechanism and emphasizes important differences between early adopters and later adopters that will be relevant to the effects of any diffusing behaviors or policies. Kristian Gleditsch and Michael Ward highlight international factors in explaining the global diffusion of democratic political institutions. In addition to domestic causes of democratization, they demonstrate that a democratic transition is more likely in a given non-democracy if neighboring countries also democratize, and "firmly reject the idea that institutional change is driven entirely by domestic processes and unaffected by regional and international events" (Gleditsch and Ward 2008). However, like Simmons, Dobbin, and Garrett, they do not go into great detail about the casual mechanism underlying how international variables affect democratic transitions.

Across the literature on international policy diffusion, international norms are treated as a potential explanatory variable rather than a topic to be explained, and scholars in this literature tend to present norm-based explanations for the diffusion of policies as an alternative to explanations that focus on strategic behavior. For example, Gleditsch and Ward present the argument that "norms and values…favor the development and durability of democratic rule" as an alternative to their explanation. Simmons and Elkins argue that one way that the policy choice payoffs can be altered are "ideational" and "work through the more subjective pressures of prevailing global norms." As I discuss in the next section, this contrast presents an incomplete picture of the role of international norms in explaining the widespread diffusion of a variety of policies and practice among states. Although these scholars do not attempt to explain international norms, I would argue that many of the substantive topics they explore can be better understood through the lens of international norm formation, as I discuss in greater detail below.

Within international relations theory one of the dominant approaches—frequently referred to as "rational" or "neoliberal institutionalism"—typically discusses norms as embedded within international institutions, and therefore generated along with them, frequently as a result of demand for interstate cooperation or through imposition by powerful states. This second alternative theory of norm development is similar to my argument in that both focus on strategic interaction between international actors, but the institutionalist theory is not intended to explain the formation of costly norms or norms that are not imposed by powerful states.

The types of norms discussed by institutionalists—such as those governing the flow of goods across borders – are distinct in that they contribute to or result from mutually beneficial international cooperation. Any risks associated with the norm must therefore be outweighed by the benefits of cooperation. Cooperative norms (also called conventions) can result from simple coordination dilemmas, such as a community's decision to drive on one side of the road, or the adoption of international aviation control regulations. Defection is automatically punished, and the gains from following the norm are clear. Norms may be sticky or path-dependent, and persist after the incentives that generated them change, but in general, the substantive focus is on norms that facilitate international cooperation by providing focal points, common knowledge, or by constraining or ordering preferences.

Similarly, scholars in economics and international law have argued that norms and other social conventions can develop "spontaneously" as a result of repeated interactions, and persist because they are Nash equilibria. My argument is distinct from these dominant theories in that it presents an alternative mechanism for the creation of international norms, and shows how consequential international norms can be generated unintentionally in a process that is endogenous to strategic interaction. Across various issue areas, signaling-generated norms may coexist with advocacy, norm-entrepreneurs, pressure from powerful states, and incentives for cooperation, although I emphasize the distinction to make the theoretical contribution of this paper clear.

A Signaling Theory of Norm Diffusion and Policy (In)Effectiveness

Some states provide better climates for foreign direct investment. Some states are more democratic, less corrupt, or more accountable. Some states make excellent allies and are well equipped for military cooperation. These and many more sources of variation between states

influence their desirability as international partners and their ability to attract international benefits. International actors may prefer to work with and reward some types of states. I refer to very loosely to states that possess desirable characteristics as "good" types. Whether any given state is a "good" type may be in the eye of the beholder, may change over time and across space, and may be very difficult to judge in a subjective manner. The central assumption is that international actors care about variation among benefit-seeking states, and international actors prefer to reward states with some types of characteristics over states with other characteristics.

I also assume that leaders of many states in the international system work to maximize their share of international benefits. International benefits are targeted toward states possessing desirable characteristics and withdrawn from states that are revealed not to possess them.

Information between states is asymmetric: governments possess accurate information about their own type, but other international actors can have difficulty judging whether another state is a desirable type. Thus, even when international actors prefer to interact with specified types of states, they cannot always distinguish good from bad types and, all else held equal, prefer to avoid rewarding states of uncertain value that might possess undesirable characteristics.

Benefit-seeking states with desirable characteristics are thus motivated to find ways to credibly signal their type to other international actors. If an attempted signal is successful in communicating a state's valued characteristic to international audiences, it is rewarded. Mimicry of successful signals can be one mechanism by which a new behavior spreads. The behavior becomes an international norm when benefit-giving actors believe that all 'good' governments send the signal. Under this condition, more states are motivated to adopt the signal, even those that must fake the signal. Thus, the behavior becomes expected among good types, may spread to bad types, and can diffuse widely even in the absence of explicit advocacy or pressure on states to adopt the new behavior.

Acceptance of the signal as an internationally held norm (or the shared expectation that all 'good' types do *X*) reinforces the incentives for states to continue the signaling behavior. The normalization of a signal also ties the behavior more closely to a characteristic that is valued by powerful international actors. Initially, these benefit-giving international actors may be indifferent to the signal, but once it is widely accepted as a behavior adopted by all states (of uncertain type) that possess the desirable characteristic, they are motivated to invest in the quality of the signal, making it more difficult for undesirable types to fake it. Therefore, when a

signal becomes a norm, it increases the costs for leaders who refuse to signal and simultaneously makes it riskier for undesirable types to attempt to signal.

Note that when a signal is accepted as an international norm, and international actors develop a shared belief that all 'good' types engage in a specific behavior, it does not necessarily mean that all states that engage in that behavior are necessarily good types. This is a simple logical point, but is crucial for understanding why certain behaviors diffuse so widely, and why they do not always appear to have consistent effects, as illustrated in several of the cases.

Sketch of a Model

To illustrate this general argument in somewhat more concrete terms, I outline a simple model of interaction between an incumbent government and benefit-giving international actors. The incumbent regime, *i*, can be one of two types, a good type (*G*) or a bad type (*B*), $i \in \{G,B\}$. The benefit giving foreign actors are denoted by *F*, and are the intended recipients of signals sent by the incumbent.[2] The incumbent chooses whether to adopt a policy or practice, *P*, or not. Adopting the policy or practice does not guarantee that the incumbent actually possess the underlying valued characteristics. If a government does not actually possess the underlying characteristics but wishes to gain the benefits anyway, it expends effort *C* to cheat, or fake the signal. Examples of cheating might be a government adopting processes related to central bank independence but maintaining the ability to manipulate fiscal policy for partisan purposes, or a government adopting national multiparty elections with no intention of giving up power should they lose.

The sequence of moves is as follows. Prior to the start of the game, *F* sets the level of benefits available to a government recognized as possessing a desirable characteristic. In the first stage, the type of the incumbent government is determined by chance. The probability that the incumbent is of type *G* is represented by γ, where $0 \leq \gamma \leq 1$, and the corresponding probability of *B* is $1-\gamma$. In order to focus on the decision to adopt a particular policy, I limit the model to the simplest case in which good types never fake (mimic) the signal ($M=0$), and the bad type always fakes the signal ($M=1$).

Excluding the countries that are already known to possess the desirable characteristic (if they exist), international actors' prior beliefs are that $\gamma < 1/2$. The incumbent chooses to adopt the

[2] This type could be expanded to include some types of domestic audiences.

policy (P=1) or not (P=0). All governments pay a marginal cost of adopting the policy, called a "transaction cost," denoted by Y, with $Y \geq 0$. All cheating is costly, and costs of trying to fake the signal are determined by the ease with which F can detect manipulation of the signal. Nature moves and the incumbent receives the reward or not.

Following the decision by i to adopt the signal or not, F allocates international benefits based on their updated beliefs about the incumbent's type. If policy mimicry is detected, no benefits are allocated. If mimicry is not detected, international benefits are (diffusely) allocated to the state. The probability that F finds evidence of cheating is r. If there is no manipulation of the signal (i.e. the true type of i is G), no evidence of mimicry is produced.

The international community reverts back to its prior beliefs about the incumbent's type ($\gamma < 1/2$) when beliefs are not pinned down by Bayes' rule off the equilibrium path.

Given the observed behavior of the incumbent, external actors accept the incumbent's signal as a valid indicator of type G or reject the incumbent as type B. Acceptance of signal is denoted by X =1, with X =1 indicating that cheating was discovered and the signal was rejected.

Summary of Timeline

Stage 1: The incumbent, i, determines whether to adopt policy or practice P.

Stage 2: F accepts the signal or not.

Stage 3: Payoffs are accrued.

Payoffs

International benefits are allocated to the government after they have chosen whether to adopt the policy and whether international actors accept the signal (i.e. whether they detect cheating). The amounts of international benefits tied to particular characteristics are exogenous, and are denoted by $A \geq 0$. They are based on the relative value of a country's characteristics to international actors. The payoff to the incumbent government is:

$$\begin{cases} A - Y \; if \; P = 1 \text{ and } X = 1 \\ -Y \; if \; P = 1 \text{ and } X = 0; \\ 0 \text{ otherwise.} \end{cases}$$

International actors are better off when they accurately support countries that have valued characteristics, and withhold support from states that adopt policies as insincere signals. They

gain *V*, when they accurately reward true types and avoid rewarding fakes. I assume that $V > 0$ when any international benefits tied to *P* exist. Thus, the payoff to international actors is:

$$\begin{cases} V \text{ if } X = 1 \text{ and } i = G; \\ V \text{ if } X = 0 \text{ and } i = B \\ \quad 0 \text{ otherwise.} \end{cases}$$

The belief by *F*, the international actors, that a government that adopts a behavior or policy *P* is a good type is updated based on the incumbent government's behavior following Bayes' rule. *F*'s updated post-election beliefs about the incumbent's type are informed by whether the incumbent adopted *P*, whether *M* was detected (with probability *r*), and *F*'s prior beliefs about whether the incumbent is a good type.

This setup can be used to illustrate that if the international benefits *A* are sufficiently high, international actors should develop the belief that all good types (but not only good types) adopt the policy or practice *P*. Such a belief means that any incumbent who does not adopt *P* must be a bad type, thus increasing the incentives for bad types to attempt to mimic the signal, even if the probability that they will be caught doing so is greater than zero. The degree to which a particular policy or practice spreads is determined by the size of international benefits associated with a particular characteristic (or signal), the probability that manipulation of the signal is detected, the cost of adopting the signal (which may differ by type), and *F*'s existing beliefs.

Discussion

For the time being, the important conclusions from the above model are that increased international benefits tied to a particular underling characteristic can lead to three related changes:

1) The shared belief among international actors that all 'good' types adopt a specified policy or practice as a signal of this characteristic.
2) The shared belief about the behavior of 'good' types causes a subsequent increase in incentives for bad types to fake the policy or practice.
3) The quality of the signal over time changes as a function of whether states can mimic the signal as well as innovation in detecting cheaters.

Although they are not modeled (for now), these three propositions yield several implications in relation to the creation of signaling norms.

First, signaling norms are only likely to diffuse among bad types if the signaling or monitoring regime is imperfect. In other words, if it is not possible for bad types to mimic the signal, the policy or practice will not diffuse, and will yield a separating equilibrium. Note again that such an equilibrium is unlikely when the possibility for mimicry exists and there are some benefits associated with the characteristic linked to the signal.

Second, signaling norms are more likely to be consequential (or informative in the long term) if the cost of the signal to bad types can be increased. Both international actors who allocate international benefits and good types of states have the incentive to increase the cost of the signal when mimicry is possible, which can create a "race to the top" dynamic. This may be true in elections (in the long term), various forms of corporate or trade governance such as forestry certification (Cashore, Auld, and Newsom 2004; Cashore et al. 2007), and international treaties.

Third, signaling norms may create global demand for a new signal of the valued characteristic if mimicry is possible and increasing the cost of the signal to bad types is difficult (i.e. if the probability that bad types will be caught mimicking the signal is very low). This may be true in the case of bilateral investment treaties.

Finally, and perhaps most interestingly, under specified conditions, signaling norms are likely to have heterogenous effects. If a policy or practice is adopted by a government that is sincerely committed to engage in the deeper reforms signaled by that policy or practice, then the behavioral change should have its intended effects. If, in contrast, a bad type adopts a policy only to attempt to mimic the signal, then the behavioral change is unlikely to have its intended effects.

Such a dynamic is only possible when the policy or practice *P* can be mimicked. The shared international expectations are most likely to have longer-term effects (and perhaps cause unintended policy changes even among bad types) when the cost of the signal can be escalated over time as bad types become better at mimicking the signal. Under these conditions, bad types become more constrained, and face a difficult choice between revealing with certainty that they are a bad type (that they are not a democracy, or that they do not have a business friendly

investment climate, for example) or accepting increased risk that they will be caught faking the signal, in which case they will have also paid the cost of adopting the new policy.

Figure 1: Three Possible Scenarios

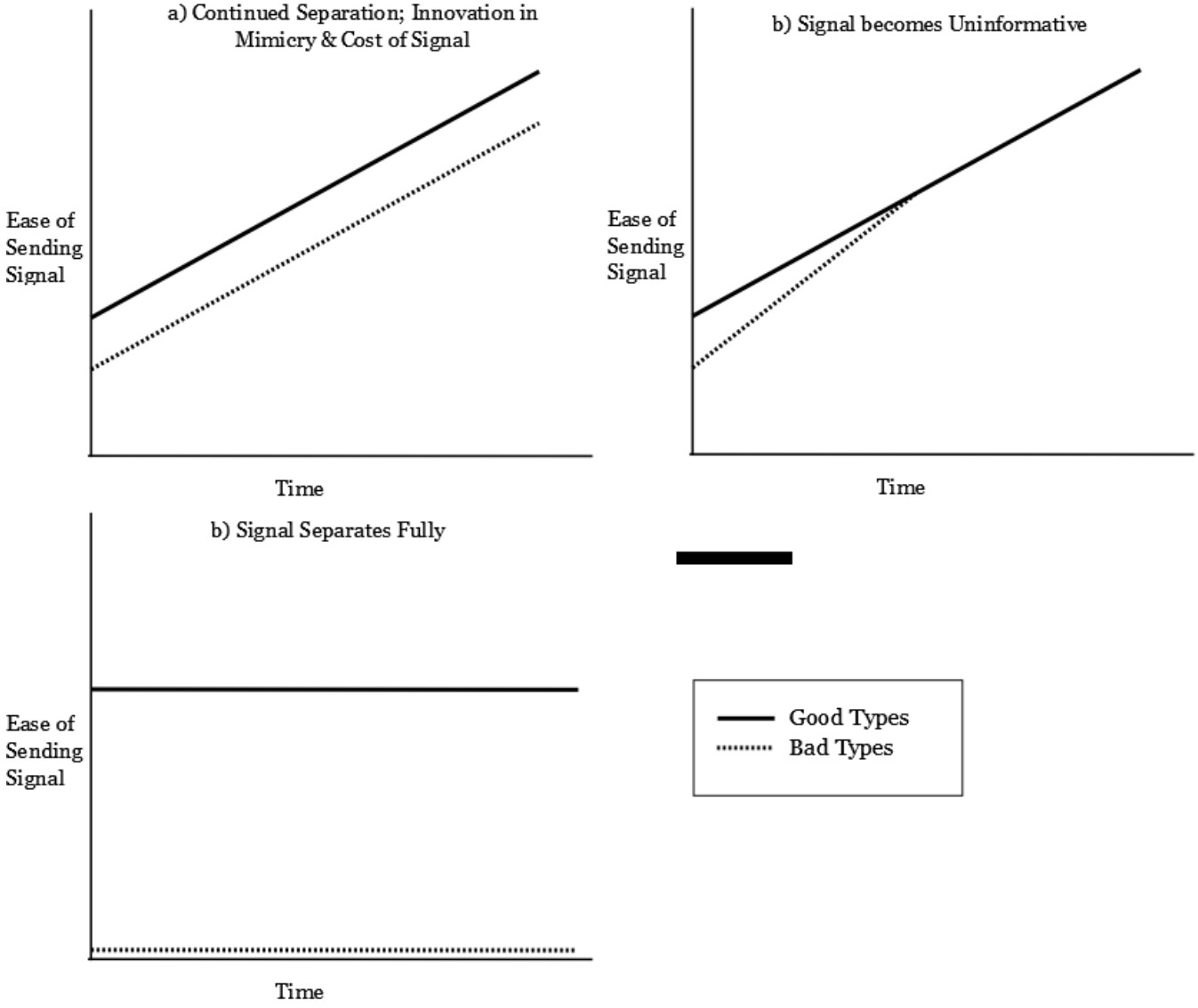


Figure 1 illustrates three possible stylized scenarios based on the ease with which 'good' and 'bad' types can adopt the signal. The specific functional forms could vary across issue areas. In Scenario A, it is always more costly/difficult for the bad type to successfully adopt the signal than the good type. The bad types may develop better methods for mimicking the signal over time, but either the recipients of the signal or good types successfully increase the cost of mimicking the signal for bad types. In Scenario B, there is no increase in the cost of the signal

for bad types, and the signal eventually becomes uninformative. In Scenario C, mimicry of the signal is not possible, and the signal is always a reliable indicator of a state's type.

Thus far this theory has been presented on a very abstract level. In the next section I use several preliminary cases to illustrate these dynamics.

Illustrative Cases

As a preliminary exercise, I explore the theory in the context of several cases currently covered in either the norm formation or policy diffusion literature.

Signaling Democracy: Elections and Competition

Democracy is one characteristic of sovereign states that benefit-giving international actors may value. The value of democracy relative to other characteristics fluctuates over time. The possibility of international benefits available to states perceived as democracies may help explain the international diffusion of many policies and practices. It may also help explain why many states have adopted formal democratic institutions but have failed to become democratic.

Multi-party elections and democracy go hand in hand, and elections are widely viewed as necessary for democracy. It is also widely recognized that elections are not sufficient for democracy. As of 2011, 95% of the independent states in the world have adopted the institution of national elections, having held at least one election for national office within six years. There are only six independent states (excluding micro-states) that did not hold any national election in the first decade of the 20th century: China, Eritrea, Libya, Qatar, Saudi Arabia, Somalia, and the United Arab Emirates. Some of the countries that held elections do not allow multiparty competition, but even that is becoming increasingly rare, as shown in Figures 3 and 4.

It is entirely plausible that national elections were adopted in each country for purely domestic reasons, and global patterns are merely a coincidence. This paper does not rule out this alternative explanation, but instead attempts to lay out one possible alternative.

National elections may have been adopted by nearly all countries in the world for the following reasons. First, democracy is perceived as a characteristic valued by powerful international actors who maintain the ability to allocate international benefits. Second, elections are believed to be a necessary condition for a democracy; a country cannot be a democracy without regular elections for national office. Third, holding elections is not sufficient for democracy, and there are many methods that governments have employed to hold elections without risking their hold on power while still maintaining the ability claim democratic

legitimacy (Schedler 2002a, 2002b). To put this in different language, the signal of holding elections to signal that a country is democratizing can be mimicked by governments that are not at all democratic, and have no intention of allowing genuine political liberalization.

Together, these reasons suggest that holding elections (and allowing minimal electoral competition) has become an internationally shared expectation for all states wishing to be perceived as democratic. Many states hold elections because they are genuinely democratic or democratizing. Other states may hold elections because they want to look like and be treated as states that are genuinely democratic or democratizing. Because it is sometimes hard to draw the line between these two groups of states, and because failing to hold elections at all is an unambiguous signal that a state is not a democracy, there are strong incentives for most leaders to hold elections and allow competition.

The international benefits tied to democracy fluctuate over time, but a significant increase occurred with the end of the Cold War. Many states, like Zambia or Togo, which adopted multiparty elections for the first time in the early 1990s did so because of international pressure. Many leaders held elections very reluctantly, and attempted to hold multiparty elections without risking their hold on power.

Manipulated elections are not new (Hermet, Rouquie, and Rose 1978), and election fraud and other tactics of manipulation have often been present in elections throughout the world. However, some tactics of manipulation are more visible than others. In terms of the signaling model outlined above, it is possible to mimic the signal of elections in order to comply with the international expectation that all democratizing countries held multiparty elections. Some governments may "fake" the signal in that they were not necessarily committed to democratization, nor willing to give up power should they lose. At the same time, international actors are aware of this possibility, and efforts have been made to make it more difficult for the signal to be mimicked. Because of continued innovation on both sides, this case is most likely an example of Scenario A described above, in which sending the signal continues to be possible for 'bad' types, but mimicry continues to be possible. There are several potentially interesting implications of this form of international diffusion. First, the signal (of holding elections with competition) spreads widely in part because it can be faked, but remains useful because it is still more costly for 'bad' types to adopt relative to 'good' types, and therefore still provides some separation.

Second, the overtime maintenance of Scenario A may require continued innovation on the part of the 'bad' types in order to better fake the signal without actually democratizing, and continued efforts on the part of the signal recipients or 'good' types to avoid mimicry and Scenario B.

There are other signals of democracy not discussed here, including inviting international election observers (Hyde 2011a, 2011b) and adopting parliamentary gender quotas (Bush 2011), and it is not clear whether the various signals of democracy are substitutes or compliments.

Figure 2: Annual Number of Elections Over Time

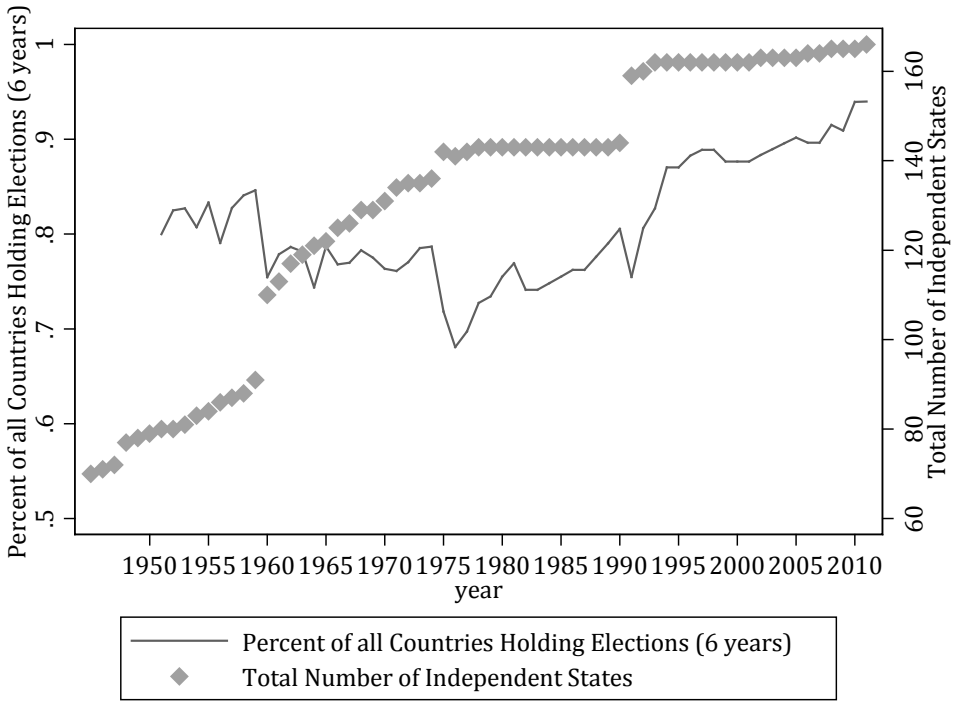


Figure 3: Annual rate of elections with and without opposition allowed (5 year moving average)

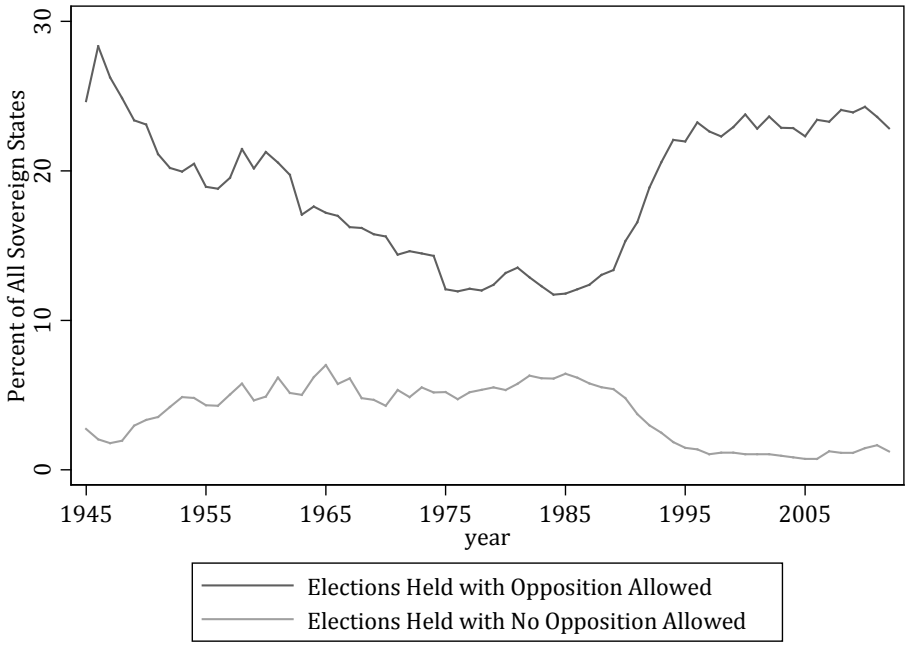
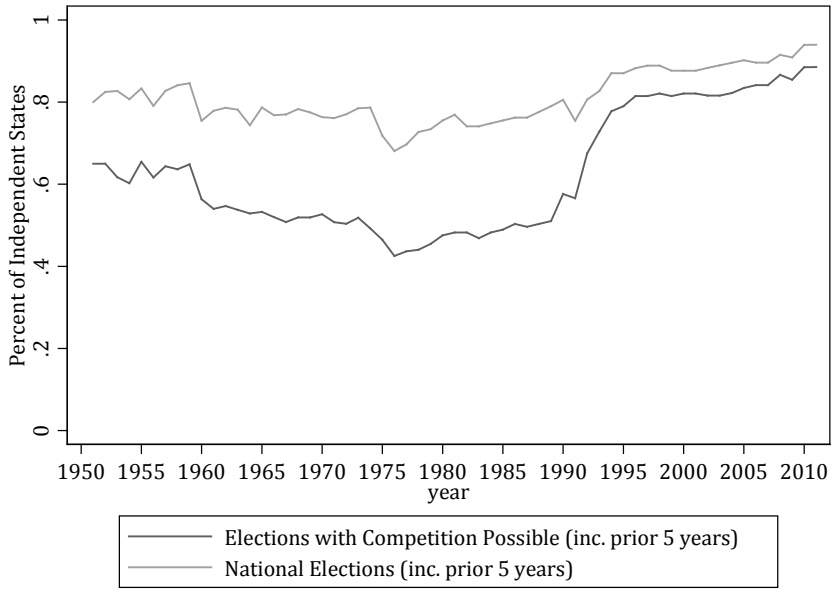


Figure 4: Percent of Independent States with Electoral Competition Possible



Signaling to Private Investors: Bilateral Investment Treaties

Like the global diffusion of elections, bilateral investment treaties (BITS) have grown from a non-existent phenomenon in the late 1950s to a widely practiced phenomenon throughout the world. Bilateral investment treaties are "agreements establishing the terms and conditions for private investment by nationals and companies of one country in the jurisdiction of another" (Elkins, Guzman, and Simmons 2008).

For governments that do not already possess well established property rights protections for foreign investors, BITs are now expected by multinational corporations as a signal that foreign investments in the country will be protected. As Elkins, Guzman, and Simmons write, before BITs existed, for governments seeking foreign investment, the existing system of customary international law "did not allow potential hosts voluntarily to signal their intent to contract in good faith" (Elkins, Guzman, and Simmons 2008, 221). Similarly, as Büthe and Milner argue with respect to preferential trade agreements and BITs,

> A government can make a more credible commitment regarding present and future economic policies by entering into international agreements that commit its country to the liberal economic policies that are seen as desirable by foreign investors (Büthe and Milner 2008, 720).

Because BITs possess an international enforcement mechanism, the treaties are arguably less costly for states whose commitment to property rights protections is genuine, and therefore represent a credible signal of a government's commitment to the property rights of investors. A number of powerful states including the US initially opposed BITs. Yet despite their opposition, many host governments embraced BITs as a method to attract foreign direct investment. Although scholars do not typically refer to BITs as an international norm, I would argue that signing BITs became a shared "expectation of appropriate behavior" for developing countries hoping to attract certain types of foreign direct investment. BITs represent a neglected international norm in the international relations literature, and an example of a signaling-based norm. The definition of international norms can easily be applied: under the norm of bilateral investment treaties, foreign investors now share the expectation that governments which desire foreign direct investment and intend to respect property right will sign BITs. Additionally, foreign investors expect that only those countries which do not intend to respect the rights of

foreign investors refuse BITs, thus generating pressure on less-desirable types of investment-seeking countries to sign BITs in order to lure investors.

Elkins, Guzman, and Simmons explain the diffusion of BITs through a competitive process that closely parallels the argument offered here, although they do not use of the term international norm nor try to explain international norm formation. According to their argument, BITs were initiated and spread precisely because they represented a credible signal of a government's commitment to enforce property rights protections for foreign investors. Those countries most in need of FDI, most likely to lose investment to competitors, and that did not already have an excellent reputation in the eyes of foreign investors (what they refer to as "inherent credibility") were the most likely to sign on to such treaties. Additionally, like the increase in democracy-contingent benefits which triggers my theory, they predict that BITs should spread rapidly when there is an increase in the global amount of capital seeking foreign investment opportunities.

My theory of norm formation also tentatively reconciles two divergent findings in the political economy literature which explain the diffusion of BITs and their effects on levels of FDI. Several scholars argue that BITs diffused because they are more costly for governments which will not respect property rights, and signing a BIT represents a credible signal of a government's commitment to respect property rights (Büthe and Milner 2008; Elkins, Guzman, and Simmons 2008). Yet other scholars, such as Susan Rose-Ackerman and Jennifer Tobin, have found that BITs only increase FDI for countries that already have a stable business environment, and little effect on low- and middle-income countries (Rose-Ackerman and Tobin 2005). If BITs in fact signal a credible commitment of respect for property rights, why are they not associated with increased FDI for all governments?

The answer to this puzzle is suggested by the dynamics of my argument. Foreign direct investors have difficulty judging whether a given government will respect property rights. All else held equal, investors prefer countries where the risk of property rights violations are low. However, once BITs were identified by investors as a signal that the government was committing itself to respect investor's property rights, refusing to sign a BIT became a signal that a government would not respect property rights. BITs diffused widely, even to countries where the business environment was less than desirable. The signaling dynamics between investors and investment-seeking governments created pressure towards a pooling equilibrium in which all

governments that *might* respect property rights were expected to offer BITs to investors. In the competitive market for FDI, many investors began to assume that governments that did not offer BITs were undesirable places to invest, and also began to believe that many less-desirable countries also offered BITs.

This overtime dynamic eventually generated pressure on governments to find additional signals of the quality of their investment climates, and offers an explanation for Rose-Ackerman and Tobin's paradoxical finding that BITs do not increase FDI to low and middle income countries (which I assume are less likely to be also be able to send other costly signals of the quality of their investment climate) (Rose-Ackerman and Tobin 2005).

Signaling Good Governance: Central Bank Independence

Independent central banks combined with transparent political institutions have spread relatively widely throughout the world and are generally interpreted as one method by which governments can commit to a low-inflationary monetary policy (Bernhard, Broz, and Clark 2002; Bernhard and Leblang 2002; Broz 2002; Franzese 1999; Keefer and Stasavage 2002). Kathleen McNamara and Sylvia Maxfield have (separately) argued that adopting "central bank independence is one way of signalling to investors a government is truly 'modern', ready to carry out extensive reforms to provide a setting conducive to business." (Maxfield 1998; McNamara 2002). McNamara criticizes literature that explains the diffusion of independent central banks as a credible commitment device by highlighting that governments sometimes adopt central banks when they do not necessarily need the policy credibility (McNamara 2002). She also demonstrates that central banks have not necessarily been successful –as the most extreme functionalist argument of central bank independence would predict—at ameliorating inflation or improving economic conditions in countries that adopt them. Alternatively, McNamara argues that central banks and other organizational structures "diffused across borders through the perceptions and actions of people seeking to replicate others' success and legitimise their own efforts at reform by borrowing rules from other settings, even if these rules are materially inappropriate to their local needs." (McNamara 2002, 48). Her argument implies that the adoption of central banks could have diffused for reasons similar to what I have argued. Like the spread of elections and electoral competition to pseudo-democratic regimes, and like the diffusion of BITs to countries that are not able to attract increase FDI, scholars have also found

that central bank independence is only associated with lower inflation under certain conditions, and that it is also subject to political and other influences that may lead to the adoption of inflationary policies (Jácome and Vázquez 2008; McNamara 2002).

Applying my theory of norm development to the diffusion of CBI would suggest that one potential reason why CBI does not always have its intended effect of reducing inflation is in part because some governments adopted the policy of independent central banks in order to appear like other states that adopted business-friendly neoliberal economic reforms. These governments may try to influence the decisions of the central bank for political gain, or appoint central bankers who will not necessarily maintain a low-inflation policy.

Other state behaviors have become signals of neoliberal economic policies, and my argument would suggest that if an important audience (in this case, either domestic constituents or international investors) develops the belief that all 'good' types of neoliberal democratic states adopt independent central banks, fixed exchange rates, capital account liberalization or other such policies, and these policies are rewarded by the relevant audiences, failing to adopt these policies begins to signal that a given government is necessarily a bad type, and not committed to neoliberal or pro-investment policies. If international actors believe that all states that have good investment climates send such signals, the behaviors can be usefully understood as international norms.

Discussion and Conclusion

How are existing theories related to my argument? If they do not directly contradict each other, and instead explain norm formation under varying conditions, when is each theory most likely to apply? I offer a preliminary answer to these questions in part by making the simplifying assumption that states or governments are motivated to comply with new behaviors (potential norms) because they think it is in their interest to do so. This is a conservative rationalist assumption, indicating only that governments avoid new behaviors if they believe that doing so will make them worse off. The assumption says nothing about the composition of potential benefits, which may include material gains, such as foreign aid and international investment, or nonmaterial gains, such as legitimacy or prestige.

The interesting question, in my view, is not whether adopting a new behavior is in a state's interest but rather which factors within the environment changed such that modifying existing behavior is perceived to be a better option than the status quo. The three (simplified)

theories of norm formation differ most clearly on why states are motivated to change their behavior. For cooperative norms, opportunities to institutionalize mutually beneficial cooperation are in a state's best interest because they directly benefit the norm-complying government through gains from cooperation or because they help institutionalize such gains from cooperation.

For advocacy-based norms, the desired change in behavior is often not in the state's interest without pressure from norm entrepreneurs. Norm entrepreneurs work to make compliance with the new behavior more beneficial to target states or to increase the costs for non-compliers. Thus, activists cause changes in the international environment and pressure states to adopt the new norm, changing their decision calculus in a manner that is distinct from the reason why states begin complying with cooperative norms. For signaling-based norms, as I argue here, changes in preferences among benefit-giving actors provide diffuse incentives for individual states to signal their type in order to increase their share of international benefits, triggering a dynamic process that ultimately leads to international norm formation (i.e. the shared belief among international actors that all good types adopt a specific policy or practice).

In comparing the reasons that states begin to change their behavior and comply with a potential norm, I suspect that singling norms fall between cooperative and advocacy norms. For signaling norms, the driving force for states to change their behavior is a change in the preferences of benefit-giving actors, although the change is not necessarily imposed or coerced by other international actors. The broader changes in preferences among international actors can be caused by norm advocacy, or by imposition by a powerful state, although my theory is noncommittal on this point.

Changes in preferences among international actors occur for a number of reasons and are treated in my argument as exogenous, but signaling norms may be closely related to broader changes in the normative environment or changes in great power politics. In contrast to advocacy norms in which the behavioral change is caused by pressure from activists or cooperative norms in which the behavioral change is caused by the belief that there are mutually beneficial gains from such a change, my argument is defined by states changing their behavior because it signals something to international actors about their own characteristics.

A given signal does not necessarily have advocates (although this is a point at which my signaling theory may converge with the advocacy theory in specific cases) nor does it necessarily

cause or enforce mutually beneficial cooperation. Complying states perceive the behavior to be in their interest because it is informative to international or domestic audiences or because international actors have developed the belief that all good regime types send the signal.

This distinction suggests a possible pattern in the conditions under which each theory is most likely to apply. When the formation of a new norm would facilitate or enforce mutually beneficial cooperation, international benefits are reciprocal and cooperative norms are most likely. In contrast, if complying with a new standard of behavior is not perceived to be in a state's interest, but other actors wish to bring about a specific change in the behavior of states, new international norms are most likely the work of norm entrepreneurs.

Situated between these two causal paths to international norms are signaling norms, which are likely when there are (new) potential gains for actors possessing certain characteristics and when it is difficult to judge which actors possess those characteristics. Note that the suggested relationship between existing theories does not imply that norm formation is automatic under any conditions. In the best of circumstances, the formation of new international norms remains unlikely. However, defining the varying logics by which new and consequential international norms are generated, and explaining why some diffused policies have heterogeneous effects, is a potentially valuable theoretical contribution.

References

Axelrod, Robert, and Robert O. Keohane. 1985. "Achieving Cooperation under Anarchy: Strategies and Institutions." *World Politics* 38 (1): 226–54.

Barnett, Michael N. 1998. *Dialogues in Arab Politics: Negotiations in Regional Order*. Columbia University Press.

Bernhard, William, J. Lawrence Broz, and William Roberts Clark. 2002. "The Political Economy of Monetary Institutions." *International Organization* 56 (4): 693–723.

Bernhard, William, and David Leblang. 2002. "Political Parties and Monetary Commitments." *International Organization* 56 (04): 803–30.

Broz, J. Lawrence. 2002. "Political System Transparency and Monetary Commitment Regimes." *International Organization* 56 (04): 861–87.

Bush, Sarah. 2011. "International Politics and the Spread of Quotas for Women in Legislatures." *International Organization*. http://ssrn.com/abstract=1608004.

Büthe, Tim, and Helen V. Milner. 2008. "The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements." *American Journal of Political Science* 52 (October): 741–62.

Cashore, Benjamin, Graeme Auld, Steven Bernstein, and Constance McDermott. 2007. "Can Non-State Governance Ratchet Up Global Environmental Standards? Lessons from the Forest Sector." *Review of European Community and International Environmental Law (RECIEL)* 16 (July): 158–72. doi:10.1111/j.1467-9388.2007.00560.x.

Cashore, Benjamin, Graeme Auld, and Deanna Newsom. 2004. *Governing Through Markets: Forest Certification and the Emergence of Non-State Authority*. New Haven, CT: Yale University Press.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. 2008. "Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000." In *The Global Diffusion of Markets and Democracy*, 384. Cambridge: Cambridge University Press.

Fearon, James, and A. Wendt. 2002. "Rationalism v. Constructivism: A Skeptical View." In *Handbook of International Relations*, edited by Walter Carlsnaes, Thomas Risse, and Beth A. Simmons, 52–72. London: Sage.

Finnemore, Martha, and Kathryn Sikkink. 1998. "International Norm Dynamics and Political Change." *International Organization* 52 (04): 887–917.

Franzese, Robert J. 1999. "Partially Independent Central Banks, Politically Responsive Governments, and Inflation." *American Journal of Political Science* 43 (3): 681–706.

Gleditsch, Kristian Skrede, and Michael D. Ward. 2008. "Diffusion and the Spread of Democratic Institutions." In *The Global Diffusion of Markets and Democracy*, edited by Beth A. Simmons, Frank Dobbin, and Geoffrey Garrett, 261–302. Cambridge: Cambridge University Press.

Goertz, Gary, and Paul F. Diehl. 1992. "Toward a Theory of International Norms: Some Conceptual and Measurement Issues." *Journal of Conflict Resolution* 36 (4): 634–64.

Hermet, Guy, Alain Rouquie, and Richard Rose, eds. 1978. *Elections Without Choice*. London: Macmillan.

Hyde, Susan D. 2011a. "Catch Us If You Can: Election Monitoring and International Norm Diffusion." *American Journal of Political Science* 55 (2): 356–69.

———. 2011b. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an International Norm*. Cornell University Press.

Jácome, Luis I., and Francisco Vázquez. 2008. “Is There Any Link Between Legal Central Bank Independence and Inflation? Evidence from Latin America and the Caribbean.” *European Journal of Political Economy* 24 (4): 788–801.
Katzenstein, Peter J., Robert O. Keohane, and Stephen D. Krasner. 1998. “International Organization and the Study of World Politics.” *International Organization* 52 (4): 645–85.
Keck, Margaret E, and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, N.Y: Cornell University Press.
Keefer, Philip, and David Stasavage. 2002. “Checks and Balances, Private Information, and the Credibility of Monetary Commitments.” *International Organization* 56 (4): 751–74.
Kelley, Judith. 2008. “Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring.” *International Organization* 62 (02): 221–55.
Maxfield, Sylvia. 1998. *Gatekeepers of Growth*. Princeton: Princeton University Press.
McNamara, Kathleen R. 2002. “Rational Fictions: Central Bank Independence and the Social Logic of Delegation.” *West European Politics* 25 (January): 47–76.
Nadelmann, Ethan A. 1990. “Global Prohibition Regimes: The Evolution of Norms in International Society.” *International Organization* 44 (4): 479–526.
Price, Richard. 1998. “Reversing the Gun Sights: Transnational Civil Society Targets Land Mines.” *International Organization* 52 (03): 613–44.
Rose-Ackerman, Susan, and Jennifer Tobin. 2005. “Foreign Direct Investment and the Business Environment in Developing Countries: The Impact of Bilateral Investment Treaties.” *SSRN eLibrary*, May. http://ssrn.com/abstract=557121.
Schedler, Andreas. 2002a. “The Menu of Manipulation.” *Journal of Democracy* 13 (2): 36–50.
———. 2002b. “The Nested Game of Democratization by Elections.” *International Political Science Review* 23 (1): 103–22.
Simmons, Beth A., Frank Dobbin, and Geoffrey Garrett. 2006. “Introduction: The International Diffusion of Liberalism.” *International Organization* 60 (04): 781–810.
Simmons, Beth A., and Zachary Elkins. 2004. “The Globalization of Liberalization: Policy Diffusion in the International Political Economy.” *The American Political Science Review* 98 (1): 171–89.
Tannenwald, Nina. 1999. “The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use.” *International Organization* 53 (03): 433–68.
Zacher, Mark W. 2003. “The Territorial Integrity Norm: International Boundaries and the Use of Force.” *International Organization* 55 (02): 215–50.