SOCIALIZING STATES: PROMOTING HUMAN RIGHTS THROUGH INTERNATIONAL LAW (*excerpts of Chapters 1 & 9*)

Ryan Goodman & Derek Jinks

TABLE OF CONTENTS

Chapter 1. Introduction: Rethinking State Socialization and International Human Rights Law

How can the international legal regime encourage states to respect human rights? There are clearly good reasons to think that like-minded states, at times, coordinate their response to common problems through international law. But, are there good reasons to contemplate a more ambitious role for the international legal order? Might the international regime, under certain conditions, encourage meaningful changes in state behavior or in the very definition of state preferences? For example, how might global institutions encourage universal acceptance of a state responsibility to provide public education for all children? And how might officials across the world be motivated to forego torture as a policy instrument even during states of emergency? In short, we want to know how exactly the international legal regime can help to promote the adoption of more humane forms of governance.

These are central questions for students and practitioners of international law and politics. The design of an effective international legal regime requires an understanding of the levers of influence on states (and other relevant actors). That is, regime design choices in international law turn on an empirical understanding of (1) the composition and functioning of social forces that influence states at the global level—for example, how exactly material payoffs, reasoned arguments, or international status might motivate recalcitrant states; and (2) the conditions under which the different modes of influence are more or less effective.

Addressing those subjects also leads to more fine-grained questions. For example, how might different modes of influence supplement or undermine

one another? What lasting effects do different modes of influence have in domestic politics following the adoption of a global norm? What is the value of partial compliance over time?

In this book, we identify three specific mechanisms for influencing state practice: material inducement, persuasion, and acculturation. We also describe the distinct, and sometimes competing, logic of each mechanism. Most importantly, we maintain that acculturation is an overlooked, conceptually distinct social process through which state behavior is influenced. The description we offer of the international legal regime and the policy recommendations issuing from this description defy conventional wisdom in human rights scholarship.[1]

A. The Empirical Study of International Law

The increasing exchange between international relations theory and international law illuminates some difficulties involved in regime design and offers useful insights to resolve them. The existing scholarship, however, also has significant limitations. Inspired by the theoretical frameworks and empirical findings of international relations research, legal scholars began to develop empirically-oriented legal analyses of international human rights law. This groundbreaking "first generation" of empirical international legal studies demonstrates that international law "matters." Nevertheless, the existing literature does not adequately account for the regime design implications of this line of research. Regime design debates often turn on unexamined or undefended empirical assumptions about foundational matters such as the conditions under which external pressure can influence state behavior, which social or political forces are potentially effective, and the relationship between state preferences and material and ideational structure at the global level. Moreover, prevailing approaches to these problems are predicated on a thin and underspecified conception of the social processes that affect states.[2]

What is needed is a "second generation" of empirical international legal studies aimed at clarifying the social mechanisms for influencing state

---

[1] This project builds on, and extends, our previous work on the topic. See Goodman and Jinks 2009; Goodman and Jinks 2008; Goodman and Jinks 2005; Goodman and Jinks 2004; Goodman and Jinks 2003.

[2] Schaffer and Ginsburg 2012: 1 ("The theoretical debate over whether international law matters is a stale one. What matters now is the study of the conditions under which international law is formed and has effects."). The debate over whether international law matters is related to the debate over whether international law is law at all. See O'Connell 2008 (summarizing the debate over the legal status of international law.

4

practice. Several recent scholarly works plainly constitute the beginnings of this next phase of empirical international legal studies.[3] This second generation, in our view, should build on this work by generating concrete, empirically falsifiable propositions about the role of the international legal regime in transforming state preferences and behaviors. Indeed, mechanism-based theorizing moves beyond mere empirical correlations and develops accounts of the properties and processes that are causally responsible for changes in actors' beliefs, preferences, and practices. A mechanism-based approach would involve systematic evaluation of whether an observed relationship reflects true causation. Accordingly, researchers would examine, for example, whether a regular pattern of outputs (e.g., similar constitutional interpretations) when actors (e.g., judges/members of the legal community) are exposed to particular inputs (e.g., foreign constitutional developments among prestigious countries) suggests the dominance of particular mechanisms of influence. This form of analysis is also conducive to evaluating the structure and content of predicted changes in practice when actors are driven by one mechanism versus another. And greater precision in identifying the internal logic of different mechanisms can uncover unintended consequences—e.g., negative interactions between mechanisms such as displacement and crowding out effects. In short, a well-specified analysis of causal mechanisms can potentially inform and fundamentally reshape ongoing debates about human rights law and practice.

First-generation scholarship has provided an indispensable but plainly incomplete framework in identifying mechanisms for influencing states. Prevailing approaches suggest that the international regime alters human rights practices either by *materially inducing* states (and individuals) or by *persuading* states (and individuals) of the validity and legitimacy of human rights law. In our view, the former approach fails to grasp the complexity of the social environment within which states act, and the latter fails to account for many ways in which the diffusion of social and legal norms occurs. Indeed, a rich cluster of empirical studies from different academic disciplines document particular processes that socialize states in the absence of material inducement or persuasion. These studies conclude that the power of social influence can be harnessed even if: (1) collective action problems and political constraints that inhibit effective material inducements are not overcome and (2) the complete internalization sought through persuasion is not achieved. We contend that this scholarship now requires a reexamination of the empirical foundations of the global human rights regime.

---

[3] Schaffer and Ginsburg 2012; Hafner-Burton 2012.

This book provides a more complete conceptual framework by identifying a third mechanism by which international institutions might change state behavior—what we call *acculturation*. By acculturation, we mean the general process by which actors adopt the beliefs and behavioral patterns of the surrounding culture. More specifically, this mechanism induces behavioral changes through pressures to assimilate—some imposed by other actors and some imposed by the self. Acculturation encompasses a number of micro-processes including mimicry, identification, and status maximization. The touchstone of this mechanism is that identification with a reference group generates varying degrees of cognitive and social pressures to conform. We do not suggest that international legal scholarship has completely failed to identify aspects of this process. Rather, we maintain that the mechanism is underemphasized, insufficiently specified, and poorly understood, and that it is often conflated or confused with other constructivist mechanisms such as persuasion. Differentiating the mechanism of acculturation and specifying the micro-processes through which it operates are profoundly important, however, for addressing questions pertaining to the adoption of international legal norms. Indeed, each of the three mechanisms—material inducement, persuasion, and acculturation—is likely to have distinct implications along a number of dimensions including the durability of norm adherence, patterns of adoption and modes of contestation.

B. Objectives of the Project

Two descriptive concerns motivate our project. First we are concerned about issues of compliance. We are thus keen to explicate how—and under what conditions—state actors are induced to obey international law and to bring their practices into line with international human rights standards. Second, we are equally concerned with the effects of the international human rights regime on state practices more broadly. In other words, we are also keen to explore the ways in which the international human rights regime might produce—intended and unintended—beliefs about state responsibilities, conceptions of illegitimate state conduct, and changes in governance structures beyond mere rule-adherence.

In our descriptive analyses, we aim to identify, differentiate, and fully articulate the mechanisms that drive state behavior. Toward this end, we emphasize the mechanism of acculturation because it is often neglected or poorly understood. Our ultimate goal, though, is an *integrated model* of the human rights regime's influence on states. Such a model would account for all the mechanisms of influence, their interactions, and the conditions under

which one or another mechanism is most likely either to effectuate change or to enhance the prospect that another mechanism will do so.

Additionally, a general normative objective also motivates the project. Our aim is to improve the understanding of how norms operate in international society with a view to improving the capacity of global institutions to promote human rights.[4] That is, our objective is to help actors to exploit mechanisms of social influence in designing and operating the global human rights regime.

In particular, our attention to acculturation can be exploited in two respects—to promote desirable norms and to arrest undesirable ones. First, acculturation can be substantially responsible for the diffusion of desirable policies across the world. Accordingly, actors and institutions can learn how to harness the mechanism of acculturation to promote human rights norms within different states. For example, the UN Office of the High Commissioner for Human Rights can learn how to spread desirable national policies by tapping into public officials' search for standardized models of modern statehood and into their concerns about national prestige.

Second, acculturation can be substantially responsible for the diffusion of undesirable policies. The substantive norm could be intrinsically undesirable—such as the spread of eugenics in the interwar period. Or the substantive norm might be generally beneficial, but the acculturation can lead to inefficient forms of policy adoption—for example, truth commissions that are not well tailored to national needs. In the case of deleterious norms, actors and institutions that have a better understanding of acculturation can learn how to overcome or impede the causal process.

Aside from the specific focus on acculturation, our analysis of the composition and interaction of different mechanisms can also improve regime design and operation. Indeed, greater conceptual clarity about the prospect and limits of each mechanism can improve decisions involving the allocation of limited resources to promote human rights. And, every step toward the development of an integrated model furthers this objective.

---

[4] The general formulation that is central in this project is how international law, as such, promotes changes in the behavior or preferences of states. It is important to note, though, that the changes sought are often only a reaffirmation of, or more meaningful compliance with, normative commitments already formally incorporated into state law and policy. See Burke-White and Slaughter 2006:350 ("International rules and institutions will and should be designed as a set of spurs and checks on domestic political actors to ensure that they do what they should be doing anyway, that is, what they have already committed to do in their domestic constitutions and laws."). Indeed, we document the widespread acceptance of international human rights standards in national constitutions in Chapter 4.

We accordingly consider in detail how a close analysis of the characteristics and functioning of each mechanism matters for regime design and operation. We link each of the three mechanisms of social influence to specific regime characteristics—identifying several ways in which an acculturation-centered approach would differ from the more common regime design approaches of material inducement and persuasion. In short, we reverse engineer structural regime design principles from the salient characteristics of underlying social processes. Through a systematic evaluation of four formal design problems—membership rules, precision of obligations, enforcement, and domestic implementation—we elaborate an alternative way to conceive of regime design. Our analysis not only recommends reexamination of policy debates in human rights law; it also provides a conceptual framework within which the costs and benefits of various design principles and advocacy strategies might be assessed. Applications relate to formal and informal aspects of the contemporary human rights regime such as: peer review among states in restricted membership organizations; the drafting of treaty text; legal practices regulating the incorporation of treaties in federal systems; and transnational advocacy groups' advocacy of economic and social rights. We maintain that the analyses and recommendations issuing from understanding the distinct role of acculturation defy conventional wisdom in human rights scholarship. Without this understanding, several characteristics of international society, in fact, will persistently frustrate efforts to obtain compliance with human rights law solely by materially inducing and persuading recalcitrant actors.

Many will sensibly argue that the best approach to the design and operation of the human rights regime would incorporate elements of all three mechanisms. This argument reflects the view that the identified mechanisms reinforce each other through a dynamic relationship that is sacrificed when a regime emphasizes one mechanism to the exclusion of others. This is an important point. However, the kind of analysis contemplated by this line of criticism (i.e., the development of an integrated theory of regime design accounting for each mechanism) first requires, in our view, identification and clear differentiation of these mechanisms. This conceptual clarification is a first step, which enables subsequent work aimed at identifying the conditions under which each of the mechanisms would predominate—and potentially reinforce or frustrate the operation of the others. Moreover, we think it useful to link specific mechanisms to concrete regime design problems. Doing so illustrates the design features suggested by each and further clarifies the conceptual commitments of each mechanism. Our analysis of regime design problems yields three models of human rights regimes—one built on each of

the mechanisms. But we do not suggest that any regime does or should exhibit all of the features of a single mechanism. The fundamental point is that, although we emphasize acculturation in what follows, we do not claim that acculturation is the most effective or most important mechanism for influencing states. Our claim is that theories of human rights law's influence--and the approaches to international human rights regime design derived from such theories--must account for the acculturative effects of the international legal regime.

C. Theorizing State Socialization

Although much of the book addresses macro-level phenomena and does so through a detailed analysis of macro-level evidence, we ultimately seek to develop a theory of the micro-foundations of global order--one that takes culture seriously and one that takes both social structure and agency seriously. This is why we emphasize organizational studies and sociological institutionalism on the one hand and the behavioral and psychological foundations of social action and social influence on the other. A fuller accounting of the mechanisms by which the human rights legal regime influences states is a necessary step in building the sort of theory of law's influence that we have in mind. Law and legal institutions are, in part, a tool for the socialization of relevant actors. Moreover, this socialization occurs by way of multiple discrete mechanisms including acculturation. These mechanisms, in turn, operate through various micro-processes at the individual level. An integrated theory of the influence mechanisms promises to facilitate better description and design of the international human rights regime.

Making a convincing case for our model of influence mechanisms obviously requires making a convincing case for global-level acculturation of states. To do so, we must prove, as an empirical matter, that this socialization process occurs at the global level and that it influences state policies and practices. We must also make a minimally plausible normative case for acculturation as an organizing principle of international human rights regimes. We make the case in full for acculturation--including its role in a broader theory of international law--in the balance of the book. Before turning to the heart of the argument, we should make clear the theoretical foundations of our project (the present section) and the specific ways in which our argument might contribute to the academic and policy literatures on international human rights law and international law more generally (Section D).

Our trifurcation of influence mechanisms--and particularly our conception of acculturation--build on two more fundamental ideas. The first idea is that

any sufficiently fulsome theory of human motivation must account for the pervasive effects of culture, social structure, and human cognition on the preference formation of individual actors. The social and behavioral sciences have developed an increasingly nuanced conception of the human actor--one that importantly qualifies both the oversimplified model of actors as wealth maximizers and the idealized conception of actors as rational, deliberative agents. This more nuanced understanding of human actors emerges from important insights in sociology,[5] anthropology,[6] behavioral economics,[7] experimental economics,[8] social psychology,[9] and brain science/cognitive studies.[10] These insights have informed regulatory reform agendas in various fields of law and policy including choice architecture;[11] corporate governance;[12] development and antipoverty policies;[13] entrepreneurship and innovation;[14] and criminal law.[15] We provide a more fine-grained explication of some of these social influence literatures in our theoretical elaboration of the three mechanisms in Chapter 2. For now, it suffices to underscore that it is increasingly well established that individual actors are influenced via various axes we characterize as acculturation. That is, human actors are embedded in a wider institutional environment--a cultural, social structural, and cognitive context--that influences preference formation and ultimately human action in multiple ways. Individual actors are influenced to change their preferences and to pursue particular courses of action by social and cognitive forces that involve neither material inducement nor persuasion.

The more difficult question, empirically and theoretically, is whether state-level policies and practices are influenced by cultural or cognitive forces substantially organized outside the state. As already mentioned, we make this case systematically in Chapters 2-4. We emphasize here, though, the second foundational idea for our project. This is the idea that states may be usefully analyzed as formal organizations embedded in, and structured by, a wider institutional environment. This idea, which grows out of sociological

---

[5] Greenwood, et al. 2008; Dimaggio and Powell 1991.
[6] Boyd and Richerson 2005.
[7] Gintis 2009; Camerer, et al. 2003.
[8] Bowles and Gintis 2011; Gintis, et al. 2006; Henrich, et al. 2004.
[9] Haidt 2012; Kahneman 2011; Kahneman and Tversky 2000.
[10] Churchland 2011; Mikhail 2011; Glimcher 2010.
[11] Thaler and Sunstein, 2009.
[12] Stout and Blair 2001.
[13] Banerjee and Mullainathan 2008.
[14] Benkler 2009.
[15] Braman, Kahan, and Hoffman 2010.

institutionalism, emphasizes the ways in which state behavior and state identity are influenced by exogenous social forces. States are formal organizations and these organizations are, in turn, part of and reflect a wider social order. In particular, we utilize institutional theories of organizations to explain some otherwise puzzling features of states.

Late-nineteenth- and twentieth-century sociological thought was, in no small measure, preoccupied with developing a general theory of formal organizations. Despite widely varying theoretical approaches, Emile Durkheim, Max Weber, Talcott Parsons, and Michel Foucault all sought to explain the organizational features of social life. The central problems for any such theory are: (1) how to explain organizational structure, and (2) how to understand the relationship between organizations and their environments. Formal organizations are formal organizations are simply tools, then organizational structure will reflect task demands conditioned by commonplace: corporations, schools, hospitals, civic associations, and, of course, governments. Traditional approaches emphasize functional explanations of these units—that is, organizations are understood (to put it crudely for the moment) as tools fashioned to address some collective problem. And if formal organizations are simply tools, then organizational structure will reflect task demands conditioned by the material impediments and resources extant in the relevant organizational environments. Organizations, on this view, are "the structural expression of rational action."[16] Accordingly, "environments," in functionalist accounts, present material, technical challenges and opportunities.

Since the mid-twentieth century, sociologists have substantially qualified traditional functionalist accounts. These contemporary approaches understand organizations as products of "institutions." The concept of "institution" is a general one referring to any regulative or cognitive feature of an organizational environment such as rules, laws, norms, and cognitive frames. The transformative insight of these institutional approaches to organizations was that formal organizations are, over time, "infuse[d] with value beyond the technical requirements of the task at hand."[17] Once socially defined institutional environments are in place, changes in organizational form are frequently driven more by considerations of legitimacy than by concern for rational adaptation or efficiency.[18] Institutions, on this view, guide interaction by providing frames or sets of meanings to interpret the behavior of the self

---

[16] Selznick 1948: 25.
[17] Selznick 1957: 17.
[18] Scott 2007; Dimaggio and Powell 1991; Meyer and Rowan 1977.

and others. Institutions thus structure the field of possible action and the ways in which organizations inherit and satisfy specific expectations. "Institutionalization" is the process by which these rules and shared meanings move from abstractions to specific expectations and, in turn, to "taken for granted" frames.[19]

The important point is that organizations are, in important respects, enactors of institutional models derived from cultural processes. States are, of course, organizations embedded in complex, global fields of action. And it is unsurprising that strands of neo-institutionalist thought have developed theories of the state and international politics. One such approach—"world polity institutionalism"—has generated substantial empirical work emphasizing the cultural and associational aspects of international politics.[20] The notion of state-level acculturation in our model grows out of the theoretical contention that states, as formal organizations, are defined by and legitimated through these fields of action. As we detail in Chapters 3 and 4, world polity institutionalism strongly supports this idea by providing a sound theoretical and empirical basis for three related propositions. First, several distinctive properties of the state are constructed by cultural processes. Second, these cultural processes are substantially organized at the global level. Third, specific features of world society accelerate the diffusion of global scripts under certain conditions. Starting with these propositions, we elaborate a detailed model of state-level acculturation. In doing so, we clarify its general character, the micro-processes that drive it at the individual level, its empirical footprint, and the evidence suggesting it is an important part of the international human rights regime (and the international legal order more generally).

Careful readers might wonder what we mean, though, by state socialization or state acculturation. As we discuss more fully in Chapter 3, our theory of state-level acculturation does not suggest that states, as such, are acculturated. We do not claim that the state is worthy of the same ontological status as a person. The acculturative forces in question are substantially organized outside the state. Moreover, these forces are manifested in state-level outcomes--in that they are reflected most commonly in formal state policy or practices. These forces are also often directed against the state. We need not contend, though, that states, as such, are materially induced, persuaded, or acculturated. Instead, we claim that patterns of formal state practice suggest that global-level institutions systematically influence state-

---

[19] Jepperson 1991: 143.
[20] Krucken and Drori 2010; Meyer, et al. 1997

level legal and policy choices. In effect, the international legal regime causes changes in state policy and practice. This macro-macro causal link, on our view, is ultimately explicable at the micro-level--even if it is most clearly documented at the macro-level. The specific causal pathway might be usefully summarized as follows. Macro-level developments influence relevant actors within states including government officials, policy advisors, members of the national and local media, issue-specific activists, and even ordinary citizens. These actors, in turn, influence the national-level legal and policy outcomes. In other words, we postulate a macro-micro-macro causal explanation.

D. Advancing the Understanding of State Socialization

The current project yields some important insights about the capacity of international society to influence states' human rights practices. Our project improves our understanding of state socialization in four principal ways. First, we introduce acculturation as a significant social mechanism for influencing state behavior. Second, we explicate the microprocesses that compose each of the primary mechanisms of social influence. We delve deeply, for example, into the psychological research on persuasion to explain how exactly it works—theoretically and empirically—to change beliefs and behavior. Third, we derive important principles for the design and operation of the international human rights regime on the basis of our empirical analysis of causal mechanisms. And, finally, we take significant steps toward the development of an integrated model of human rights regime design – one that, by definition, accounts for all three mechanisms of social influence, the conditions under which they are likely to be effective, and the potential for them to supplement or counteract one another.

We also hope that the following pages will advance the understanding of state socialization in other respects as well. Some of these contributions deserve highlighting here. Moving from descriptive to normative claims, they include the following:

> 1. We theorize a "check list" of quantitative and qualitative indicators that suggest when acculturation (rather than persuasion or material inducement) explains the diffusion of a particular norm.[21]

---

[21] Such a list is an important supplement to debates about the proper way to conceive of the empirical markers of acculturation. And, indeed, we qualify contentions that we made in

2. We analyze an enormous body of empirical research, from a range of different disciplines, indicating the power of acculturation in global politics. We examine findings that show different degrees of explanatory power for acculturation. That is, in studies of different areas of state policy, acculturation is either (i) a significant source of international influence; (ii) the dominant source of international influence; or (iii) the only source of international influence. For the purposes of our larger argument, however, we are satisfied with even the most modest of those empirical claims.[22]

3. We analyze some of the latest social network research on the interaction of states in international affairs. This line of research provides an important supplement to previous work on the role of networks in international law.[23] We also refine some of the existing social network analysis by examining which mechanisms might explain the spread of ideas and preferences across a network structure. The book thereby brings acculturation to social networks analysis.

4. We address normative concerns about exploiting acculturation to promote desired policy outcomes. We directly address concerns about hegemony and other distributional consequences for the international order. Notably, analysis of the network structure of intergovernmental and nongovernmental organizations at the global and regional level helps inform this analysis.

5. We examine some unintended consequences of diffusion via acculturation that can pose critical challenges to human rights law. For example, we discuss the prospect of "races to the middle"—whereby states that would otherwise aspire to heightened levels of human rights protection gravitate toward lower expectations or toward the standardization of mediocre definitions of success. Indeed, in some areas of human rights law one might need to consider whether the aggregate result is desirable—that is, whether the benefit of raising the

---

earlier writing—for example, whether the twin findings of cross-national isomorphism and domestic decoupling constitute sufficient evidence of acculturation.

[22] In this part of our analysis, we also show that the leading empirical study on international human rights law (Simmons 2009) supports our theory of the significance of acculturation—despite the author's contention that the data generally support a different school of thought.

[23] See, e.g., Slaughter 2004; Burke-White 2004; Raustiala 2002.

14

performance of a lower rung of states (measured by their human rights violations) outweighs the cost of reducing the performance of higher rung states.

6. We show counterintuitive effects of some design choices in the international legal regime. For example, we argue that legal precision is not conducive to compliance or to building an international rule-of-law under some conditions. And we show how peer-review systems within intergovernmental organizations can backfire if not properly matched to the organization's membership rules.

7. We directly address the criticism that acculturation-led social change may result in only partial or superficial compliance with human rights law.[24] We examine national-level sociopolitical consequences of changes in state practice due to acculturation.[25] We explore, for example, effects on domestic political opportunity structure and mobilization of social groups that favor human rights improvements.

8. We examine in detail how different mechanisms of social influence might "crowd out" one another under certain conditions.[26] We borrow from economic and psychological research on such crowding out effects in interpersonal relations, and postulate a research agenda for studying such effects on the international level.

These points, we submit, advance our understanding of how international legal regimes influence states. They also suggest several ways in which our project might contribute to the further development of world polity

---

[24] To formulate the criticism and its implications with precision, we develop a typology of various forms of decoupling. Sociological institutionalism predicts persistent decoupling—the disconnect between form and function-- in any social field. As we develop fully in Chapters 3-4, the general tendency is well documented--as is its prevalence in state practice--but the literature has not systematically differentiated the varieties of decoupling.

[25] In this analysis, we build on research that closely studies the relationship between (a) the *motivating force* behind a state's initial decision to adopt an international standard and (b) subsequent domestic *institutionalization or retrenchment* with regard to that adopted practice over time (Weber, Klaus, Davis, and Lounsbury 2009; Zelner, Henisz, and Holburn 2009). Other studies, for example, have separately examined the decision to commit and the decision to comply (Simmons 2009; Hathaway 2002; Hathaway 2003).

[26] We primarily examine how material incentives crowd out other social mechanisms; however, we also explore crowding out effects between persuasion and acculturation -- two constructivist mechanisms -- and between different types of acculturation.

institutionalism (and sociological institutionalism more generally) as well as international relations theory. Our project supplements world polity institutionalism by grounding that work in a mechanism-based analysis. Such an approach facilitates the identification of concrete, testable propositions about whether, and under what conditions, cultural context shapes actor preferences or behavior. The detailed empirical "check list" that we develop demonstrates the payoff of our overall approach in this regard. We also identify several micro-processes by which world cultural effects might be manifested at the individual level.[27] Furthermore, we analyze a number of interaction effects that might facilitate a better understanding of whether, and under what conditions, we should expect normative "de-institutionalization." In addition, we build on earlier work of others,[28] in an effort to integrate world polity research and social network analysis.[29] Finally, we place significant emphasis on relationships between the international order and domestic politics, especially after the adoption of a globally prescribed norm.

Our project supplements international relations theory in political science in several respects as well. Once again, the mechanism-based approach— coupled with the detailed empirical "check list"—aids in developing concrete, testable empirical predictions. The lack of sufficiently precise, testable predictions has, of course, been an important problem for the constructivist research agenda in international relations theory. It is particularly useful, in our view, to develop a theory of mechanisms and micro-processes driving top-down social constructivism. Sociological institutionalism, and specifically world polity institutionalism, avoids the circularity problem endemic to some constructivist research. Constructivist research often fails to distinguish adequately between explanatory and outcome variables--often claiming that both variables are mutually constitutive. The "new institutionalism" in sociology distinguishes between "organizations" and "institutions"—the concept of "organization" refers to the formal apparatus (and its purposes) whereas the concept of "institution" refers to all regulative and cognitive

---

[27] In addition, our typology of different forms of decoupling can facilitate greater theoretical and empirical precision in describing how broader institutional contexts might be reconciled with the particular identities of relevant actors and organizations that enact a legitimated script.

[28] Beckfield 2010; Beckfield 2003.

[29] For example, several world polity studies examine whether a state's increased membership in intergovernmental organizations (IGOs) is associated with the adoption of a norm. However, these studies do not consider the characteristics of other states in the networks formed by IGO membership and the relative position of a state in the network. If states acculturate due to increased ties to IGOs, we would expect the composition (and network relationships) of other states in the IGOs to be an important factor.

features of the organizational environment such as rules or shared beliefs.[30] The important point is that our approach avoids circularity problems by clearly differentiating explanatory variables (institutions) and outcome variables (organizations).

E. Outline of the Book

In Part I of the book--Chapters 2-4--we develop our theory of social influence. Chapter 2 identifies and distinguishes the three mechanisms of social influence as well as the micro-processes of each mechanism. Our emphasis throughout this chapter is the fundamental behavioral and social structural logic of each mechanism. In Chapter 3, we elaborate the theory of state acculturation. We explain our claim--common in international relations research in political science and world polity research in sociology--that socialization occurs at the state-level. We also identify a series of theoretical and empirical claims suggesting the presence of global-level acculturation--developing a detailed "check list" describing with some precision the empirical footprint of acculturation. In Chapter 4, we survey the available empirical research suggesting state acculturation generally as well as the research suggesting its importance in the realm of human rights. In Part II of the book, Chapters 5-7, we apply the theory to several specific regime design problems--analyzing whether specific mechanisms and their attendant micro-processes imply any specific design characteristics for international human rights regimes. Chapter 5 addresses membership rules. Chapter 6 analyzes the scope and content of legal obligations. And Chapter 7 considers monitoring, enforcement, and implementation. Part III of the book addresses objections and outlines how our project fits into larger research agendas in international legal studies. In Chapter 8, we address the important criticism that human rights regimes ought never be substantially grounded in acculturation because this mechanism, at best, produces shallow reform. We outline several causal pathways by which acculturation can produce meaningful, enduring change.

---

[30] Of course, many "institutions" can also be understood as "organizations" depending on the object of the study. For example, in a study of the organizational features of hospitals, the state (including perhaps most prominently, regulatory agencies) are part of the institutional environment within which hospitals operate. But, in a study of the organizational features of state regulatory agencies, it is the agencies themselves that are analyzed as "organizations" (and "institutions" in this study would include the salient features of the wider cultural environment in which the agencies are embedded). The important point is that our approach avoids circularity problems by clearly differentiating, as an analytic matter, explanatory (institutions) and outcome variables (organizations).

Chapter 9 outlines the ultimate objective of our project--an integrated theory of the international legal regime's influence on state human rights practices. We summarize several important features of such a theory. Chapter 10 summarizes the overall argument and offers some thoughts on productive directions for future research.

[. . . .]

Chapter 9. Toward an Integrated Model of State Socialization

The primary objectives of this book are to provide a more comprehensive typology of the mechanisms of global social influence, to analyze the drivers and characteristics of each mechanism, and to demonstrate the significance of the analysis for institutional design. More fundamentally, the objective is to provide a framework within which future research in international legal scholarship—empirical, theoretical, and doctrinal—might more fully understand the role and capacity of international law in promoting human rights. The ultimate objective is, then, to develop an empirically grounded model of regime design that integrates all three mechanisms. Such an integrated model must account for the distinct qualities of, and interactions between, the processes of material inducement, persuasion, and acculturation. Although the specification of an integrated model is well beyond the scope of this book, we discuss some general features of such a model in this Chapter. We identify four features of a well-designed integrated model: (1) take seriously acculturation as a practical, conceptually distinct mechanism of social influence; (2) consider interaction effects; (3) consider sequencing effects; and (4) specify the conditions under which the various mechanisms would be most, and least, effective. We discuss each in turn.

A. Taking Acculturation Seriously

An integrated model should seriously consider acculturation processes. Indeed, acculturation has been systematically undervalued and, at times, misunderstood in debates about human rights regimes. As discussed earlier, commentators rarely invoke acculturation; when they do, it is often conflated with persuasion or unexplained. As we discuss throughout the book, acculturation provides a conceptually distinct, empirically supported, and potentially effective approach to promoting human rights through international law. We offer three further points to supplement these more fundamental claims.

First, there is good reason to suspect that material inducement and persuasion will prove ineffective or glaringly inadequate in the human rights arena. States generally lack sufficient resources, interest or willpower to sustain an effective strategy of material enforcement or material inducements more generally. This unfortunate, but widely understood, feature of international politics figures prominently in material, rationalist-based theories of human rights law—often highlighted as a reason why human rights law is doomed to failure.[31] Persuasion-based approaches, on the other hand, account and aim only for complete internalization—and there is little evidence to suggest that this is a reliable method of socializing bad actors. Indeed, theories of persuasion do not provide a useful way to think about partial or incomplete internalization. And the circumstances in which actors will be convinced to improve human rights by appeals to their value system will necessarily be limited. In short, the prevailing approaches will often prove ineffective or insufficient.

Second, acculturation is potentially highly effective—enabling broad, robust, and enduring influence over the trajectory of state human rights practices. As we point out in Chapter 8, acculturation frequently generates deep reforms, and frequently generates shallow reforms that evolve into deeper reforms over time. We also addressed a potential conceptual confusion—the worry that readers might confuse the distinctive empirical footprint of acculturation (decoupling) as the necessary result of acculturation. Building on that analysis, here we briefly address the more straightforward worry that acculturation is unlikely to produce important change because cognitive and social costs are too low and too diffuse to exert any meaningful influence on relevant actors. The fundamental point here is that this kind of objection is based on a flawed assumption—the notion that cognitive and social interests are valued according to the same metric as material interests. This assumption suggests that acculturation effectuates change only when "competing" material interests are trivial. Cognitive and social interests associated with acculturation are, however, generally incommensurable with material payoffs associated with material inducement. A similar mistake is to suppose that state actors will weigh the costs of social "sanctions" or the benefits of social standing against material payoffs.[32] We maintain that this ultimately flawed view, even in its most sophisticated form, derives from a narrow conception of culture, in general, and emulation, in particular—one that presumes actors ultimately abide by cultural norms primarily for material

---

[31] *See*, *e.g.*, Goldsmith and Posner 2005.
[32] See, e.g, Guzman 2008.

19

reasons.[33]

Contrary to that conception, we argue that material costs are often not weighed against cognitive benefits—and actors often hoard social legitimacy for its own sake. The processes of identify formation and identity maintenance are guided by cognitive frameworks that are not subject to rational cost-benefit analysis. Similarly, material costs may not be weighed against social benefits. Social preferences are often incalculable, as a practical matter. And actors often engage in materially costly, high risk, and self-destructive practices to avoid social disapproval or to maintain self-respect. In Chapter 2, we discuss such matters at a conceptual level when contrasting social and material sanctions. In Chapter 3, we examined the issue empirically by analyzing how state actors make decisions in pursuit of status and cognitive comfort despite grave material sacrifices. Recall, for example, the findings presented in studies of the "costs of isomorphism." States experience massive economic losses and resource deficiencies by enacting global scripts that encourage building and sustaining administrative ministries,[34] and by following Western modes of scientific inquiry.[35] States also forsake national security interests (recall China's embrace of arms control[36]) even when the very survival of the country is at stake (recall Britain's abnegating use of chemical weapons to defend itself against a potential German invasion[37]). Those findings are consistent with other studies demonstrating that leaders wage wars to avoid loss of face or to control sovereign territory deemed essential to the identity of a nation-state.[38] Additionally, even if there is a frontier beyond which tradeoffs between material and social interests are calculable, actors may decide not to deviate from legitimated modes of behavior due to a set of competing priorities. That is, for relevant actors in many contexts, it will be more important to maintain a sense of belonging in a global or regional community than to avoid material costs associated with a particular course of action. The important point is that individual and organizational actors are influenced by powerful acculturative pressures regardless of the material incentives to act otherwise—and regardless of the calculable cognitive and social payoffs.

Third, acculturation is extraordinarily important to regime design because other mechanisms may often be predicated, in a meaningful sense, on its

---

[33] See Johnston 2007 (describing mimicry by socialized states as a practice in pursuit of survival); Zucker (1987 or 1977); [cites].

[34] [cite]

[35] [cite]

[36] See Johnston 2007; see also supra __.

[37] See Goodman and Jinks 2005: 1771-75; see also supra __.

[38] See, e.g., O'Neill 2001; Goodman 2006: 107.

effectiveness. Both material inducement and persuasion presume a foundation of generally shared beliefs, values, and cognitive frames. Material inducement, for example, requires shared methods and common conceptions to administer material force most effectively. Community-wide understandings are needed to deter targeted actors (e.g., to communicate desired human rights practices and to communicate the likelihood of punishment for failure to comply), to preclude subjective and self-serving interpretations of violations, and to establish proportionate penalties.

Acculturation may also serve as the cultural predicate for acts of persuasion—it sets the overarching frames that actors invoke to convince audiences to respect human rights. Indeed, frames are a product of common cognitive orientations and shared cultural assumptions.[39] "[F]raming efforts can be thought of as acts of cultural appropriation … seeking to tap highly resonant ideational strains in mainstream society."[40] One might object that human rights frames are relatively fixed such that acculturation's value in providing the foundation for persuasion is more a historical fact than one of ongoing or future significance. The cultural system, however, is more fluid than that objection presumes. For example, globally institutionalized understandings of what it means to be a "modern" state (a part of modernity itself) changes over time. The scope of particular rights—e.g., the meaning of equality, the core of economic and social rights, the content of legitimate security interests—changes considerably in response to cultural and social shifts. In short, acculturation is an ever vital precondition for the optimal exercise of other mechanisms. As a result, fully understanding and exercising material inducement and persuasion requires fully understanding and exercising acculturation.

As a final note, we should add that acculturation often complements the other mechanisms even after these conditions for success are in place. First, acculturation can increase the use of material inducements. For example, acculturative forces can motivate actors to expend their own resources to punish norm violators. As we discuss in Chapter 8, acculturation can generate pervasive third-party punishment even when such punishment contradicts the material interests of the third party.

Second, acculturation can bolster the effectiveness of both negative and positive material inducements. Consider first negative inducements. As acknowledged by proponents of approaches that rely on material force, a sanctions regime is more durable and productive if the role assumed by

---

[39] McCammon, Muse, Newman and Terrell 2007: 725; Gamson and Meyer 1996: 275.
[40] McAdam 1994: 36, 37-38 (emphasis omitted).

punishers is highly legitimated.[41] State actors may thus resist or retaliate against enforcement measures by states or institutions that lack legitimacy or that originate from outside a legitimated community. Acculturation can also amplify the social influence of positive material inducements. Consider EU expansion and the work of scholars— who favor material explanations of the human rights reforms that have been adopted by candidate states—scholars such as Judith Kelley[42] and Frank Schimmelfennig and his colleagues[43]. According to Kelley's research, the decision to place human rights conditions on accession to the EU (using the carrot of access to the European market) in conjunction with persuasion and social pressure yielded the highest overall effectiveness in encouraging East European states to adopt reforms respecting ethnic minorities.[44] As Schimmelfennig and his collaborators acknowledge, a cultural commitment to "return to Europe" may have been a "necessary condition of rule adoption if we restrict it to political elites: there were no cases of rule adoption without strong identification of the governing elite with Europe and the West."[45] In contrast, elites who depended on the political support of nationalistic groups adverse to West European identity (e.g., Slovakia under Vladimir Meciar) acted differently. They did not make concessions—despite the economic advantages of EU membership. Indeed, state actors may be willing to forego substantial material benefits that require membership in a group with which they do not identify. Finally, material incentives can initially inspire behavioral changes, and acculturative pressures that are linked to political membership may help sustain those changes over time. Admittedly the available evidence is indirect and speculative. That said, consider Kelley's acknowledgment of a striking pattern that is difficult to explain with material inducements alone: "In this study, there are actually no examples of reversal in the policies" protecting ethnic minorities after the state obtained EU membership and material incentives were then lifted.[46] The EU apparently succeeded in employing material inducements to gradually admit new states into a dense institutional environment in which conformity with minority rights protection was a behavioral norm de-linked from direct

---

[41] See, e.g., Ikenberry 2001: 30-31, 52-53 (discussing importance of legitimacy in maintaining (post war) power-based global order); Buchanan and Keohane 2006: 405; see also Chapter 9.B (discussing retaliation by targets of punishment).

[42] Kelley 2004.

[43] See, e.g., Schimmelfennig, Engert and Knobel 2006; Schimmelfennig and Sedelmeier 2004.

[44] Kelley 2004

[45] Schimmelfennig, Engert and Knobel 2006: 29, 236-37.

[46] Kelley 2004: 425, 449.

material incentives. This unusual scheme and the special nature of social identity associated with the EU may thus avert negative interactions between mechanisms that we describe below.[47]

Third, acculturation will often further the cause of persuasion. As discussed in Chapter 8, acculturation differentially empowers and legitimates actors (individuals and organizations) committed to the cause of human rights—giving these actors voice in the domestic political process. This political empowerment increases the structural opportunities of these actors to persuade other relevant domestic actors and institutions to embrace international human rights norms. Also, this social legitimacy will also often increase public trust of these actors, facilitating efforts by these actors to persuade recalcitrant audiences. Once again, acculturation has the potential to increase the number and effectiveness of persuasive encounters. In these various respects, acculturation can perform an important function in strengthening the effect of other mechanisms to promote human rights.

B. Negative Interactions between Mechanisms

An integrated model should consider potential negative interactions among the three mechanisms. Simply put, deploying one mechanism can undermine the effectiveness of another. Some of these dynamics are already well understood at a high level of abstraction. For example, human rights scholars recognize that economic sanctions can produce a "backlash" among target actors thus undermining other efforts at social change.[48] And policymakers often try to consider such effects. A well-specified analysis of the mechanisms of social influence, however, can provide a more comprehensive and detailed understanding of the potential for human rights promotion strategies to countermand one another. Indeed, these dynamics are

---

[47] Note that Kelley likely underestimates the effect of acculturation. Kelley compares the effectiveness of (1) efforts at acculturation/persuasion without conditional EU membership versus (2) efforts at acculturation/persuasion with conditional EU membership. She codes EU membership solely as a material inducement. See, e.g., Kelley 2004: 37-39. However, it is difficult to disentangle material and acculturative benefits in this case. Recall, for example, that political elites in Eastern Europe experienced considerable social and cognitive benefits from (re)inclusion in the European family. Indeed, they generally already considered themselves rightfully part of the European community and fervently sought formal validation of that identity. Thus the instrument of membership may, in addition to its material power, constitute a stronger form of acculturative pressure than the milder forms of external criticisms and diffuse social pressures.

[48] Risse, Ropp and Sikkink 1999.

most apparent when one examines them at the level of mechanisms and the micro-processes that comprise the mechanisms.

Consider how strategies based on material inducements—positive or negative incentives—can become incompatible with acculturation- and persuasion-based strategies. According to the processes of acculturation, under certain conditions actors partially or completely internalize social or cognitive scripts. Over time, behavioral adherence to the script may be described as "intrinsically motivated."[49] Substantial empirical evidence demonstrates, however, that explicit material inducements (punishments *or rewards*) can "crowd out" intrinsic motivation for engaging in prescribed behavior.[50] This evidence suggests that actors are often sufficiently motivated to conform to social norms for nonmaterial reasons. Where this condition obtains, a strategy based on material incentives can have perverse consequences—producing higher aggregate levels of norm violations.

Persuasion and acculturation may similarly countermand one another. Negative interactions can occur, for example, if the former focuses attention on resolving particular substantive disagreements among states when the latter stresses abstract commonalities. Persuasion and acculturation strategies may also conflict when the former highlights the prevalence of human rights violations as a framing device and the latter, to avoid the institutionalization of undesirable behavior, casts such violations as aberrant.

Analysis of negative interactions introduces a level of complexity that may be difficult for regime designers to absorb. In some cases, the analysis may be most helpful as a post-mortem: when a regime breaks down and reformers seek explanations of ineffectiveness or perverse results. And in other cases, only one or two types of negative interactions could be relevant. A consideration of all negative interactions might be theoretically optimal. However, only consideration of a few mechanisms and under very limited conditions will be manageable. In this light, six types of interactions are worth considering.

1. Conveyance of prevalence information

Negative interactions between material inducement and acculturation may

---

[49] Frey and Jegen 2001: 589; Ostrum 2005.
[50] See generally Id. (surveying empirical studies of interactions between extrinsic incentives and intrinsic motivation); Bowles 2008: 5883; Ostrum 2005: 260;. Deci, Koestner, and Ryan 1999: 627. Cf. Eisenberger and Cameron (1996: 1153-1166); Cameron and Pierce 1994: 363-423.

occur due to implicit information conveyed through the operation of material incentives. An instrument that employs material inducement can suggest that the proscribed practice is widespread. And information about the prevalence of a practice influences the behavior of target actors in several ways. Such information might increase noncompliance where other-regarding preferences are conditioned on notions of reciprocity and fairness. If target actors are "conditional reciprocators," they may abandon their other-regarding preferences when a material inducement cues them to believe that other actors are defectors. Substantial empirical evidence documents that many pro-social actors are "conditional reciprocators" and that prevalence information weakens their otherwise robust commitment to pro-social behavior.[51]

Prevalence information might also increase noncompliance by unintentionally promoting a social norm supporting the proscribed behavior. Because actors often emulate orthodox or widespread social practices, the instrument might cue actors to form beliefs about the characteristics and behavior of other group members and, in turn, to adopt those practices for themselves.[52] That is, "conformists" alter their own practices when the instrument signals that a significantly high fraction of other actors behave in a particular way.[53] And a policy that invests heavily in material inducements to achieve rule conformity can communicate to targeted actors that their reference group (of similarly targeted actors) does not highly value self-motivated rule adherence.[54]

Some research also suggests that persuasion-based interventions trigger the same effect. For example, Deborah Prentice's work on domestic social norms demonstrates that overt efforts to persuade actors to discontinue a behavior can produce in a "boomerang effect."[55] A persuasion-based campaign can implicitly (and unintentionally) suggest that deviance from the policy objective is prevalent (and socially accepted by others). International human rights institutions as well as human rights advocacy groups are designed to document and report—even widely publicize—human rights abuses. This information gathering and information conveyance strategy increases public awareness of human rights abuses in an effort to persuade actors to change

---

[51] *See*, *e.g.*, Bowles 2008 and Gintis 2000, forthcoming 2008*; Fehr [cites]

[52] Sliwka 2007: 999-1012; Sliwka.

[53] Sliwka 2007: 1000-01. These groundbreaking empirical studies have informed analyses of domestic law and public policies from tax enforcement to environmental protection to crime control and public order maintenance. *See*, *e.g.*, Kahan, 1997*; Kahan, [cite]. Frey 1997; Frey & Stutzer 2008

[54] Frey & Jegen 2001: 594, 602-05.

[55] Prentice [cite].

their behavior. This approach arguably reinforces the sense that human rights abuses (even of the most extreme variety) are commonplace and that other priorities of good governance—such as economic growth, public order, and national security—routinely trump the protection of individual rights. In other words, the prevailing approach might similarly create a "boomerang effect" under certain conditions. Notably, Prentice has recently analyzed the application of her research for international human rights, and, along with Andrew Woods, we have as well.[56]

2. Overjustification and social signaling

Material inducement can crowd out socially motivated adherence to a norm through an "overjustification" effect.[57] Overjustification can occur when a particular course of action is justified by both normative sentiments and material incentives. In other words, compliance with a social norm or rule is overjustified if actors have multiple, ontologically distinct reasons to observe the norm or rule. Substantial empirical evidence demonstrates that overjustification adversely influences levels of norm compliance in several ways.

Consider first cases of overjustification and social signaling. Many actors, motivated by concern about their status in a community, will adopt prosocial behaviors to signal their moral character. Due to acculturative pressures at the global level, state actors are often keen to adopt particular practices to demonstrate their commitment to internationally legitimated human rights. The introduction of a material incentive, however, can over-justify compliance and thus degrade its value as a social signal.[58] That is, "the presence of incentives may … reduce the value of generous or civic-minded acts as a signal of one's moral character"[59] and the actor may find that "even a small material reward over-justifies his good deed."[60] As a consequence, material

---

[56] Prentice 2012; Goodman, Jinks, Woods, 2012.
[57] Bénabou and Tirole 2006: 1652.
[58] Bénabou and Tirole 2006*;* Fehr and Falk 2002: 710 ("Moral behavior is often considered to be moral for the very reason that it is undertaken despite pecuniary incentives to the contrary. Paying people for their moral behavior is, therefore, a contradiction in itself because it means that their behavior can no longer be considered as moral. For example, if you are paid for your honesty most people will no longer evaluate your honest behavior as moral behavior. Since moral behavior typically is associated with social approval, paying for moral behavior means that approval incentives will be reduced."); Ariely, Bracha and Meier forthcoming 2008.
[59] Bowles and Hwang 2008: 1811, 1813.
[60] Bowles 2008*:* 1609.

incentives weaken the commitment of these actors to the pro-social norm.

This overjustification effect would also impede signals sent in the other direction—from the social group to the individual actor. Overjustification compromises the ability of actors to identify and interpret signals from the social environment about appropriate conduct. That is, it would be difficult for actors to draw important inferences from the behavioral patterns of states—to discern, for example, whether governments with good human rights records are acting out of a principled belief about how states ought to behave or out of an instrumental calculation of material payoffs. For instance, such noise would make it difficult for the group to signal that a true consensus exists that modern states reject torture as normatively abhorrent. An alternative message is that self-regarding states are inspired to avoid material penalties for engaging in torture—or to obtain material rewards for eschewing torture.

Of course, the existence of an instrument employing material inducements itself sends a signal that the community condemns the proscribed behavior and is willing to invest resources and exert force against bad actors. Hence, the instrument and each instance of material inducement can signal strong social support for the (human rights) norm. And, the *absence* of a material inducement-based strategy might send the opposite signal—that the community lacks strong social support for the (human rights) norm. This "expressive function" of the material inducement-based strategy is an important countervailing effect, to be sure.[61]

Our point is not to deny or in any way discredit that bit of conventional wisdom, but rather to underscore several important ways in which it must be qualified. Most importantly, our analysis suggests that overjustification can weaken social system-actor signals as well as actor-actor signals—which cautions against jointly employing material and social incentive strategies in all cases. Whether the "expressive function" of punishment swamps this overjustification effect will turn on numerous considerations. For example, the expressive function is compromised if penalties and rewards issue from actors with insufficient social standing vis-à-vis the signaled actors—a narrow band of donor countries, a remote foreign court, unrepresentative segments of civil society, a hostile country. The expressive function is also diminished or lost if material payoffs result simply from structural conditions rather than from a purposefully directed (and publicly endorsed) system of incentives. More fundamentally, the expressive function of punishment might work only when the proscribed behavior is broadly, unequivocally, and manifestly understood

---

[61] See, e.g., Feinberg 1965: 397; see also Sunstein 1996: 2021; Cleveland 2001: 89-90.

as inappropriate.[62] In other words, the expressive function of employing material inducements might predominate only when there is little risk of overjustification given the well understood status of the norm in question.

3. Overjustification and self-perception

Overjustification triggered by the provision of material incentives might also adversely affect targeted actors' self-perception. Material inducements might interfere with the very cognitive processes that lead to internalization of a norm. Overjustification causes some actors to lose cognitive track of their motives for abiding by a norm and to attribute their actions to material incentives: "Individuals sometimes do not understand their own motives perfectly …. If monetary incentives are set for an activity, then an individual concludes that it performs this activity because of those incentives. If the incentives are abandoned, motivation is reduced as compared to a situation where there never have been extrinsic incentives."[63] As a result, actors otherwise inclined to observe a social norm because they considered observance an extension of their identity or required by their internal value system, might instead perform the act only because and for so long as the balance of material incentives weighs in favor of compliance. Moreover, these psychological processes help explain why actors who have already internalized a prosocial norm will often not act on it once material incentives are introduced. Actors motivated by concerns about their self-worth "consider themselves as less praise-worthy when they collect money [for engaging in moral behavior], which reduces the psychological incentive to perform the activity."[64]

4. Overjustification and self-determination

Finally, overjustification reduces the perception of self-determination in target actors—decreasing long term compliance with the overjustified norm. Much empirical evidence suggests that the provision of material incentives often compromises individuals' perception of self-determination and thus degrades their intrinsic motivations for engaging in a behavior.[65] "When

---

[62] Consider the weak to nonexistent "expressive" effect associated with the criminalization of regulatory offenses and/or minor crimes. [cite]

[63] Sliwka 2003: 2-3; Lepper, Keavney and Drake 1996: 32, 24-25; Lepper and Greene 1978: 109.

[64] *See* Fehr and Falk 2002: 687, 710-11.

[65] Bowles 2008.

people perceive an external intervention as a restriction to act autonomously, intrinsic motivation is substituted by this external intervention. The locus of control shifts from inside to outside the person. The person in question no longer feels responsible but makes the outside intervention responsible instead. However, this shift in the locus of control only takes place when the intervention is considered to be controlling."[66] The crowding out effect can be dramatic. Several studies demonstrate that when actors would otherwise seek to engage in a practice "in the absence of other rewards, the introduction of explicit incentives may 'overjustify' the activity and reduce the individual's sense of autonomy,"[67] thus driving down the aggregate levels of prosocial behavior.[68]

Although this research clearly demonstrates that the provision of extrinsic incentives decreases intrinsic motivation, the relevance of this finding for our project is less clear. The complication is that, as a conceptual matter, neither moving part in the self-determination research program maps perfectly onto the three mechanisms. The concept of "intrinsic motivation" includes socialized end states generated by both persuasion and some acculturation-based strategies—namely, those that involve cognitive pressure. Hence, the crowding out effects identified in the self-determination studies tradeoff with intrinsic motivations brought about by either persuasion or deep acculturation. The notion of "extrinsic incentives" includes material inducement and some acculturation-based strategies—those that involve social punishments and rewards of various forms. Indeed, self-determination research finds that social and symbolic rewards are often perceived by actors as controlling and thus degrade their intrinsic motivation in some circumstances.[69] This is a terrifically important point because it suggests that some forms of acculturation (partially internalized, social pressures) may crowd out other forms of acculturation (completely internalized, cognitive scripts). Hence, acculturation may not always be a step along an evolutionary path to complete internalization.

5. "A fine is a price"

A related research program shows how fines often release actors from concerns about social disapproval thereby increasing noncompliance with the relevant social norm. Conventional wisdom suggests that fines reduce

---

[66] Frey and Stutzer 2008: 406, 412.
[67] Bowles 2008: 1607.
[68] Frey & Jegen 2001: 594; Frey and Stutzer 2008.
[69] Deci, Koestner and Ryan 1999 but see Eisenberger and Cameron 1996.

infractions through infractions; fines are also thought to signify the social unacceptability of a behavior—thus reinforcing the societal pressure to abide by the norm. Substantial empirical evidence, however, documents the opposite behavioral effects. In certain circumstances, actors perceive "a fine is a price" simply to be paid in exchange for engaging in a proscribed behavior.[70] In field studies, subsequent to the introduction of a fine, the rate of misbehavior increased and stabilized well above pre-fine levels.[71] One explanation of this result attributes the behavioral change to information conveyance.[72] That theoretical account, however, requires essentially perfectly rational, perfectly self-interested actors;[73] and, even if accurate, would support an important point—rational actors are willing to violate a social norm by purchasing the prerogative to do so. Another explanation of this finding is that fines change the social meaning of norm violations. If actors experience social discomfort in violating a norm, the fine releases them from such pressure.[74] The fine essentially changes actors' perception of the nature of the obligation.[75] The latter explanation is consistent with other studies demonstrating that monetary incentives encourage actors to feel justified in violating a social obligation.[76]

Applied to the human rights context, state actors may be more likely to violate substantive rights and to breach procedural obligations (e.g., requirement of periodic reporting before international bodies) if they must pay for such acts monetarily. Monetary costs could include financial compensation, interstate reprisals in the form of countermeasures or restrictions on foreign assistance or trade. Such costs may become perceived as the price of doing business in the way the state actors otherwise prefer. One

---

[70]Gneezy and Rustichini 2000: 1; Fehr & Falk 2002: 711 ("giving potential norm violators the opportunity to free themselves from following a social norm by making them pay for the norm violation may backfire"); Cardenas, Stranlund and Willis. 2000: 1719; Bohnet, Frey and Huck 2006: 131.

[71] Gneezy and Rustichini 2000a: 3, 8.

[72] The fine provides new information that misbehavior will not result in more severe penalties; actors are no longer deterred by the uncertain threat of worse sanctions. See Gneezy and Rustichini 2000b: 791.

[73] Gneezy & Rustichini 2000a; 13.

[74] Gneezy and Rustichini 2000a: 14; Fehr and Falk 2002: 709 ("While in the baseline condition there was no ambiguity about the fact that being late constituted a violation of the rules the imposition of a price conveyed the message that the commodity of "being late" could now be bought.")

[75] Gneezy and Rustichini 2000a: 14; Fehr & Falk 2002: 711 ("the introduction of the fine not only reduces the disapproval for being late but parents also no longer consider being late as blame-worthy.")

[76] [cites]

commentator's insightful analysis of the potential effects of monetary compensation on the rights of indigenous peoples illustrates the concern. She explains: "Incorporating 'just compensation' as a liability constraint on government limitations of aboriginal rights could create a risk that 'justice' will be equated with compensation. Once compensation is provided, it may lead to a practical extinguishment of aboriginal claims in the minds of regulators and the public."[77]

C. Sequencing Effects

An integrated model should also consider various "sequencing" effects. That is, an integrated model might emphasize different mechanisms at different stages of the institutionalization of a norm. Of course there are no universal rules to apply across contexts. We can, nevertheless, derive lessons about general tendencies that should subsequently be considered in individual cases. Those context-specific decisions will be better informed by understanding factors such as the potential detrimental social effects of various sequencing choices, opportunities for mutually reinforcing mechanism interactions and the distinct functions that each mechanism can perform during different periods in the development of a human rights regime.

One general lesson involves the systematic changes in preferences that actors experience by their very participation in and exposure to an institutional environment. Because acculturation and persuasion alter state preferences over time, international organizations might incorporate more flexible administrative devices such as renegotiation clauses or encourage sunset provisions on treaty reservations—essentially devices that recognize that the preferences of states are endogenously formed by their interactions within the regime. A human rights regime might also enhance its effectiveness by demanding modest initial commitments and ratcheting up obligations over time. Strategies could include allowing supervisory organs to expand their authority incrementally and creating opportunities for optional protocols only after an organization has existed for an extended time period. Voting rules within IGOs might also anticipate greater consensus on issues over time. Regime architects could, for example, appease hesitant states by requiring larger supermajorities to bind member states later in the life of the institution. In short, mechanisms that are ineffective at one stage can be effective at another due to preference changes within institutional setting. Regime architects should build structures in anticipation of those cultural shifts.

---

[77] Metcalf 2008: 385, 446 (citing Gneezy and Rustichini, 2000a).

Our analysis throughout this book provides insight into potential social effects that result from alternative sequencing choices. When, for example, should material inducement precede or follow acculturation and persuasion approaches? One important factor is the existing strength and distribution of social expectations and cognitive beliefs supporting the normative goals of material inducement. In global affairs, the use of material force is risky because it is often either prematurely, haphazardly, or seldom employed. One precondition for effective material inducements is strongly motivated enforcers. As the example of strong reciprocity suggests,[78] material sanctions without the prior social and cognitive alignment of important actors (especially third party punishers) will be employed inconsistently, if employed at all. Premature punishment—prior to the institutionalization of a norm—can also result in a (greater) backlash by norm violators who feel unjustly penalized. In addition, as the crowding out literature suggests, attempting and then removing material incentives can undermine acculturation- and persuasion-based strategies. Indeed, some evidence suggests an "afterglow effect"[79] whereby the crowding out of intrinsic motivations endures long after the removal of material incentives.

On the other hand, employing material inducements at an earlier stage can accrue specific benefits. When long-term political and economic support for administering material force is lacking, punishments and rewards may be best reserved for limited circumstances such as: an initial phase in the development of a human rights regime or the initial point at which states join or being to participate in an organization. For example, material payoffs could incent states to join organizations in which they are later subject to measures that rely on persuasion or acculturation. As recent scholarship on China suggests, once (even powerful) states join multilateral organizations, path dependency may lead to greater levels of socialization.[80] An additional reason to emphasize material inducement earlier involves other crowding out effects. Negative interactions—Category 3 (*overjustification - self-perception*) and Category 4 (*overjustification - self-determination*)—occur when actors are already intrinsically motivated prior to the introduction of material incentives. Similarly, other negative interactions—Category 2 (*overjustification – social signaling*) and Category 5 (*fine is a price*)—occur when actors have begun to internalize societal pressures prior to the introduction of material incentives. Accordingly, foregrounding material inducements before actors have traversed

---

[78] See Chapters 8 and 9.B (discussing strong reciprocity and motivations to punish).
[79] Irlenbusch and Sliwka 2005.
[80] *See generally* Johnston *.

down the road of socialization may be more effective. That suggestion does not contradict other designs that we just discussed, but it does make their application more complicated. Awaiting the institutionalization of a norm is advisable if one is trying to minimize retaliation by actors who are punished for violating human rights, but waiting can amplify the crowding out effects among actors who would otherwise respect human rights. The important point is to consider those distributional effects and to examine such tradeoffs when possible.

As a final example of social effects of sequencing, consider benefits to emphasizing acculturation prior to other mechanisms. First, acculturation could enhance the effectiveness of material inducements. Instruments utilizing material inducements can have a "crowding in" effect when monitoring and enforcement is conducted by a peer group. Of course, a peer group is an antecedent condition for acculturation, not a product of acculturation. However, acculturation, through the communication and sharing of common practices, can help reinforce the sense of a community. A more fully acculturated group would accordingly enhance the positive effects of peer enforcement. Moreover, acculturation can help develop community-wide schema—for evaluating human rights standards, the definition of violations, and acceptable justifications—thus sharpening the framework that a system of material incentives needs to operate most effectively. Notably, these forms and benefits of delayed onset of material inducement reflect, in many respects, the evolutionary path of the European Convention on Human Rights and its member states.[81] Second, consider the benefits of deploying acculturation-based strategies prior to persuasion. Under certain conditions, a regime might concentrate on exploiting the effects of acculturation to outline and solidify broad principles of agreement, before investing heavily in persuasive techniques to define obligations more precisely. Even if acculturation were not a precondition for persuasion, it would still remain highly valuable in determining the cultural frames that actors invoke in carrying out persuasive endeavors.

D. Conditions for Mechanism Success

An integrated model should endeavor to identify the conditions under which the various mechanisms operate successfully. For instance, the

---

[81] *Cf.* Helfer and Slaughter 1997*:* 314-17 (discussing the importance of incrementalism in the evolution of European Community law and European human rights law); but cf. de Burca 2011.

effectiveness of all three mechanisms will likely vary according to the socioeconomic and political conditions within states. In particular, it is important to assess the economic capacities of states to monitor human rights practices and sanction human rights violators (as well as the capacity of violators to absorb or pass on those costs). Another factor is the composition of cultural and structural relations at the global or regional level. For example, the effectiveness of various influence strategies will turn on the structure of interstate networks, the axes along which relevant states share important characteristics (including religion, ethnicity, and language), and the distribution of military and economic power. Some of these considerations may be obvious now that we have examined the significance of each mechanism. We, accordingly, focus our attention on less obvious considerations. We group these observations under two headings: (1) targeting capacity and target actor characteristics (which should be considered together); and (2) influence agent characteristics.

1. Targeting Capacity and Target Actor Characteristics

An important consideration for institutional design involves the capacity to direct pressure against specific actors. There are two elements to this consideration. That is, targeting capacity will turn on the composition of the audiences subject to (material, persuasive, social, or cognitive) pressure and the technical power to hone the instrument to affect the intended target.

Poor targeting capacity disfavors material inducement. When the exercise of material force is imprecise, target actors are likely to pass on costs to the market or to the domestic population. Crude targeting capacity also risks economic spillover effects onto neighboring countries and trading partners. And those externalities may threaten the ability to maintain continuous, and long-term political support for such initiatives. The advent of "smart sanctions" is a salient recognition of some of these problems, and a partial corrective.

Poor targeting capacity also disfavors persuasion. There are obvious difficulties in making persuasive appeals to different kinds of audiences. That is, the content and emphasis of one frame will often be in tension with another framing strategy. Consider, for example, the different types of appeals that may be needed to convince both local rights activists and multinational financial interests to mobilize against a rights abusing government. Or consider the challenges the Chief Prosecutor of the International Criminal Court faces in attempting to obtain political support from differently oriented states—he may try to appear to some as an idealist with global ambitions and

to others as a pragmatist with economic stability of particular regions of the world in mind. Of course new information technologies compound these difficulties by making it ever more complicated to keep multiple messages and multiple audiences separated. Hence, a persuasive appeal that resonates with one audience may quickly undermine support from another group.

In contrast, poor targeting capacity generally does not hamper the effectiveness of acculturation. Targeted actors cannot "pass on" reputational and cognitive costs. And, acculturation is generally less concerned about well-honed impacts or multiple audiences. Subjecting any part of the field of actors to human rights scripts is generally beneficial. One significant caveat involves the crowding out effects identified above. If the audience includes actors who are already intrinsically motivated to abide by a human rights norm, the exposure to global social pressure may degrade their commitments. Regime architects and practitioners should accordingly weigh those distributional effects if possible.

2. Influence agent characteristics

Another important consideration is whether the hypocrisy of influence agents, especially when it is obvious and sustained, undercuts their use of different mechanisms of influence. Does a state or group of states that promotes human rights in other countries need to abide by the same standards to be fully effective? Apparently not for material inducement, because the sheer force of material incentives determines the target actors' decision calculus. Persuasion should not, in theory, depend on the internal consistency of the influence agent's practices and message either—the merit of the idea determines the target audience's response. However, persuasion is often mediated by relationships of trust and by the perceived character of the messenger. Hypocrisy may accordingly reduce an audience's receptivity to the message for reasons not specifically or directly related to the core precepts of persuasion.[82]

In contrast, hypocrisy will often directly and significantly stymie efforts at acculturation. This feature of acculturation is part of the reason why the mechanism constrains powerful countries that seek to employ soft and hard power in international affairs, as we discuss in Chapter 8. More generally, effective social pressure depends on both the legitimacy of the influence agent and the legitimacy of the human rights regime—and the legitimacy of both is

---

[82] See *supra* note __ (discussing whether such forms of influence are better classified as persuasion or acculturation).

undercut by double standards. Acculturation driven by cognitive impulses is also weakened by double standards. Sustained and blatant hypocrisy constitutes the type of exception that makes it more difficult to articulate and spread a coherent universal script. These effects are especially damaging if hypocrisy is displayed by culturally salient actors, which maybe the very actors likely to be taking the lead in promoting human rights.