The Economics of Civil Society

Naomi R. Lamoreaux, Yale University and NBER

&

John Joseph Wallis, University of Maryland and NBER

August 2012

The Economics of Civil Society

Organizations are ubiquitous in developed economies. In the United States today, the average person is likely to work for one organization, worship at another, socialize in still others, vote for candidates run by an organization called a political party, volunteer in parent-teacher organizations to improve local schools, donate to charitable organizations, and turn for advice or medical care to professionals trained and licensed by organizations. Virtually every good or service that an individual consumes is provided by an organization. Organizations are so much a part of our experience that it is easy for us to regard them as naturally occurring, and indeed one could make the case that the ability to form organizations is one of the most important characteristics that distinguish humans from other animals. But organizations do not just spontaneously appear in societies; they have a history. The kinds of organizations that humans have created, the ease with which they can be formed, and the role they play in society have changed dramatically over time.

This conference focuses on the dramatic growth that occurred with the onset of modern economic development in the number of organizations that constitute "civil society"—that is, the economic, political, religious, educational, etc., organizations that occupy an intermediate space between the organizations that structure the most personal spheres of human life (for example, families) and those that constitute the state (for example, government).[1] Although virtually all societies larger than bands of hunters and gatherers have such intermediate organizations, the literature on civil society conventionally focuses on the early nineteenth century, when the

[1] This definition is standard in the literature. See, for example, Shils 1991. Some scholars exclude from their definition organizations whose purpose is to earn profits (see Neem 2008), but we deliberately embrace the broader definition.

number of organizations in western countries suddenly mushroomed, changing the way the economy, the polity, and the larger society functioned. Most scholars take a positive view of this development, arguing that the organizations formed during this period served as schools for democracy, that they fostered an appreciation of diversity, that they were the critical elements of a new public sphere that disciplined the state. But others take a more pessimistic view, countering that many of the new organizations were exclusionary in their purpose—that they were bastions of illiberalism that reinforced racial and ethnic divisions in society and worked to limit rather than to expand democratization.[2] Scholars also disagree about the processes that drove the explosion in the number of organizations. Some explain the change as a byproduct of the religious and political conflicts of the late eighteenth century, others of urbanization and commercialization, still others of technological improvements that lowered the cost and increased the efficiency of transportation and communications (Habermas 1989, Hall 1995, Skocpol 1999).

There is much of interest in this literature, but we think it also misses much that is important. Because scholars have been preoccupied with the political, social, and cultural roles performed by these organizations, they have neglected their economic character—the extent to which the spread of organizations both stimulated, and was stimulated by, economic growth. Moreover, because their emphasis has been on the voluntary nature of these organizations, they have missed what we think is the key change of the period: the shift in government policy toward organizations. During the early nineteenth century governments in a small number of countries abandoned their previously strict control over who could form organizations and for what purposes and granted their citizens a virtually unlimited ability to form organizations of all

[2] For differing views, see the essays in Keane 1988 and Hall 1995. Compare also Putnam 1993, Rosenblum 1998, and Kaufman 2002.

types. More importantly, they offered people seeking to form organizations access to legal forms and enforcement mechanisms that enabled those organizations to be larger, more complex, and more effective. Governments did not simply refrain from repressing organizations whose formation they did not specifically sanction; they consciously and deliberately encouraged their growth and spread by making critical organizational tools widely available. As a consequence, we would argue, these governments dramatically and permanently altered the structure and functioning of their societies—so much so as to bring into being a new social equilibrium.

The purpose of this paper is to make our case for the fundamental nature of this transformation and to highlight questions that we think such an organizational perspective on history raises. It is not our intention to draw up a blueprint to structure your contributions to the proposed volume. Nor do we insist that you "buy" our interpretation. Indeed, the discussion that follows should be taken as a set of plausible hypotheses strung together with a historical narrative. Our goal is to focus your attention on the shift from limited to open access to organizations—to get you thinking about what it entailed, how it came about, and the role it might have played in sustaining both political democratization and economic development. The paper begins with a conceptual discussion that explains why, in our view, governments have traditionally found it in their interests to limit citizens' ability to form organizations and why the shift to open access was so revolutionary. The second part then uses the history of the transition in Britain and the United States to problematize conventional accounts of the development of civil society. The third suggests some possible directions for future work.

1. Theoretical Overview

1.1. What Do Organizations Do?

Organizations coordinate human activity, and in the process they create rents. The first step in articulating a theory of what organizations do, therefore, is to be clear about what we mean by the term rent. The simplest definition of an economic rent is a return above opportunity cost. If a worker is willing to work for $10 an hour (the value of the best alternative use of his time) and receives a wage of $15 an hour, the rent for an hour's work is $5. A consumer who is willing to buy a pair of shoes for $15 but pays only $10 receives a rent of $5. A producer who is willing to sell shoes for $5 but sells a pair for $10 receives a rent of $5. These examples all involve standard economic activities such as production and consumption that can easily be valued in monetary terms, but the concept of rent applies to human choices more generally. Two individuals who like each other enjoy a rent from their relationship. The rent is the subjective value they place on their relationship compared to alternative relationships they might have with other people. The rents that organizations create are very often non-monetary. They involve value that is created by forming individuals into groups.

Two features of rents play a central role in the argument that follows. First, rents are both relative and multidimensional. Suppose, as above, that a consumer values shoes at $15 a pair and can buy them from a particular producer for $10. The rent the consumer receives from buying the shoes is $5, but if he can buy an identical pair of shoes from another seller for $11, then the rent he receives from buying from the first seller in particular is only $1. Rents on such different dimensions often move in opposite directions when circumstances change. Extending the shoe example, if the number of sellers of shoes increases, the rent the consumer gets from buying shoes may increase, even if the price does not change, because he may enjoy increased

variety or increased ease of purchasing. At the same time, the rent he gets from buying from any specific seller is likely to decrease.

Second, rents create incentives to perform actions (make choices), and the probability that arrangements between people will continue in the face of uncertain and changing circumstances is directly related to the size of the rents associated with the action. If a consumer agrees to buy shoes from a producer for $10, that agreement is more likely to continue if the consumer receives $5 in rents from buying shoes than if he receives a rent of only $1. The extent to which the producer believes she can count on the consumer's continued business thus depends on her perception of the rents the latter receives. More generally, parties are more likely to make investments in relationships that continue through time when each perceives that the other obtains rents from the relationship.

Organizations are bundles of relationships. As such they create rents in two basic ways. The first is characteristic of all relationships that persist over time. When two individuals come to know each other and expect to interact in the future, they have a relationship. Relationships create rents when the alternative to which they are compared is the prospect of dealing with strangers whom one expects never to meet again.[3] These rents come both from our increased knowledge of the other person and from the expectation that our interaction will continue. These elements enable us credibly to coordinate our behavior.

This coordination is the second source of the rents that organizations create. For many activities, people who work in teams are more productive than people who work individually. If the organization is a firm that produces goods, the gains can be measured in terms of physical output. But again, the gains from coordination are not limited to standard economic activities.

---

[3] When we get to know a person we may learn that we do not want to interact with him or her, but even that negative information produces a rent in comparison to dealing with a person whom we do not know.

Churches are organizations that coordinate behavior in ways that enhance the value of the community and the religious experience. Individual church goers receive rents from their participation in the church's activities, and it is those rents and the personal knowledge of each other that results from participation that enable church goers to coordinate.[4]

Organizations, then, provide a framework for relationships that are more valuable to individuals than one-shot interactions with strangers. The value of relationships makes it possible for people to coordinate their actions, and that coordination in turn generates rents in the form of higher output or benefits than could be obtained by a comparable group of uncoordinated (unorganized) individuals.

1.2. What Holds Organizations Together?

Understanding how organizations work has been a mainstay of the new institutional economics, beginning with Ronald Coase's (1937) insights about the firm and continuing on through Oliver Williamson (1975 and 1985), Sanford Grossman and Oliver Hart (1985), and a host of others. As Robert Gibbons has argued in a series of papers designed to draw together lines of inquiry in economics and sociology, organizations can be thought of as interlaced bundles of relationships and contracts (1998, 1999, 2003). While some organizations can be described as self-enforcing sets of relationships sustained by repeated interaction and the existence of rents, most rely on some form of contractual enforcement using third-parties. A robust theory of organizations should encompass both relationships and contracts, rather than relying on one or other as the "organizing" principle.

---

[4]Organizations are not the only way that people can coordinate. The gains from specialization and division of labor can be obtained in markets, in which the price mechanism coordinates individual decisions.

One starting point for a theory of organizations is the folk-theorem intuition that two individuals can maintain a relationship over time if both individuals receive a rent from the relationship. The players in the folk theorem receive rents from their specific relationship, so their individual identity and the nature of their partnership matters. The existence of rents makes their relationship incentive compatible. The folk theorem partnership is what we call an *adherent* organization, an organization where both or all members have an interest in cooperating at every point in time (North, Wallis, and Weingast [hereafter NWW] 2009). Adherent organizations are inherently self-sustaining or self-enforcing; they do not require the intervention of anyone outside of the organization. Mancur Olson's famous "Logic of Collective Action" (1965) essentially relies on the existence of rents enjoyed by members of the organized group, which he calls selective incentives, to explain voluntary associations. Members only cooperate if the rents are positive and, critically, if the rents are only attainable within the organization.

As in the shoe examples above, the higher the rents the more predictable is the behavior of members of the organization. That is, partners can sustain a higher degree of cooperation when members of the relationship expect to receive higher rents on an ongoing basis. Members who are pushed to the margin are not reliable partners: if a member receives total benefits that are just equal to the total costs of membership, then rents are zero and that member is indifferent to cooperating. The behavior of indifferent partners is unpredictable. Any small change in circumstances may lead them to defect. Organizations want to ensure as much as possible that all members earn some positive rents so that their behavior is predictable.

If the members of an adherent organization look forward into the future and anticipate that rents may not be sufficient to ensure the cooperation of every member at every point in time, then they will expect defection and cooperation may unravel. There are, however, ways for the

members to protect against defection. For example, they may insist on hostages as insurance against the possibility that rents will become zero or negative at some point. The threat that a hostage will be killed imposes large penalties on defection, making possible incentive compatible and time consistent arrangements for the organization. The various folk theorems lay out how such punishments for deviators (non-cooperators) might be credibly imposed (Benoit and Krishna 1985, Fudenberg and Maskin1986).[5]

The folk-theorem logic can explain the existence of organizations. However, organizations that depend only on the coordinated interests of their members without recourse to external enforcement are likely to remain small. Ensuring cooperation is expensive, particularly when cooperation is attained through the continual *ex ante* transfer of real economic assets or costly threats to destroy economic assets. Take the example of cartels. Before the passage of the Sherman Antitrust Act in 1890, cartels were legal in the United States but their contracts were not enforceable in court (Freyer 1992). Firms in many industries formed them, but the cartels tended to fall apart whenever a downturn (or even the threat of a downturn) undermined the rents they earned for their members (Chandler 1977, Lamoreaux, 1985). The few that managed to achieve some longevity devoted substantial resources to detecting and punishing cheaters. For example, the Joint Economic Committee (JEC), a railroad cartel formed in 1879, had a central office that collected weekly statistics on the shipments of its members and had the authority to order members to cut prices immediately upon detecting cheating. This punishment strategy kept rents high enough for the cartel to survive (Ulen 1980a and 1980b, Porter 1983), but it seems to have been quite costly. Data collected by Thomas Ulen, who viewed the JEC as a success, indicate that incidents of cheating occurred on average in nearly one out of every three

---

[5] An historical example comes from the slave trade, where British merchants insisted that their African counterparts place relatives on slave ships in exchange for credit to use in acquiring slaves. If the African merchants failed to live up to their bargain, their relatives could be sold into slavery. See Lovejoy and Richardson 1999.

weeks between 1880 and 1886 (Ulen 1980a; see also Ellison 1994). In Germany, by contrast, cartel contracts were enforceable through the courts. As a result, not only were cartels were much more stable and enforcement costs lower, but they were able to increase the rents they earned for their members by coordinating other activities. For example, chemical firms cooperated to finance R&D, and steel firms used their organizations to smooth the flow of inputs to their plants (Chandler 1990, Fear 2005).[6]

The difference in the capacity of the German versus the American cartels resulted from the formers' ability to appeal to an external agency, a third-party, to enforce the terms of their agreements.[7] Organizations that rely on some form of external enforcement of agreements are *contractual* organizations. Anything that an adherent organization can do a contractual organization can do, but many things that contractual organizations can do are impossible to accomplish with purely adherent organizations (NWW 2009).

It is difficult to overstate the importance of contractual organizations. Those of us who live in societies with open access to organizational tools may have trouble appreciating just how many of the organizations we consider "voluntary" are contractual, not adherent, organizations. One of the authors was commissioner of a soccer league for six to nine year olds organized by a boys and girls club. The club, as a matter of course, obtained liability insurance for the commissioner. Even though the boys and girls club was completely voluntary—there was no coercion involved whatsoever—it was well understood that an aggrieved or upset parent had the ability to sue the club, the coach, and the commissioner if their child was harmed through

---

[6] German firms were able to use the IG form to increase the stability of their cartels by exchanging stock in each other's enterprises.

[7] Gabriel Kolko (1985) has argued that railroads in the U.S. recognized the value of third-party enforcement and lobbied (unsuccessfully) to secure government support for railroad cartelization in the Interstate Commerce Act of 1887. Scholars have disagreed about whether government support was necessary for the cartel's survival (see also MacAvoy 1965, Ulen 1980b, and Binder 1988), but there is no question that federal assistance would have lowered the cartel's enforcement costs.

inappropriate behavior. In other words, the larger society had provided this voluntary association with extremely sophisticated and powerful organizational tools to structure and enforce its arrangements. Virtually all organizations in modern societies are contractual in this sense, no matter how informal they appear to be. They all swim in a sea of organizational tools so pervasively present that participants often do not even notice their existence.

1.3. Third Parties and Private Ordering

Many theories of organizations assume that the institutional capacity to enforce rules and contracts already exists in the larger society, residing in something one might call a government.[8] The question then becomes how to arrange relationships in such a way as to maximize their net value and then divide the surplus among the parties. But, of course, the institutional capacity to enforce rules and contracts in the larger society did not always exist. In this section we consider the problem of how third-party enforcement arises in the absence of government—that is, how private organizations can serve effectively as enforcers for each other.

Avner Greif's discussion of "the community responsibility system" in medieval Europe offers a particularly clear example of how such a scheme might work (Greif 2006a and b; Greif, Milgrom and Weingast 1994).[9] In Greif's analysis, merchant guilds arose in cities that had important trading relationships with each other and facilitated commerce by reciprocally enforcing deals that involved each other's members. A merchant from one city (say Genoa) could confidently go do business in another city (say Hamburg), even if he had never been there

[8]For example, Bolton and Dewatripont begin their *Contract Theory* with the explicit assumption that "the benchmark contracting situation … is one between two parties who operate in market economy with a well functioning legal system. Under such a system, any contract the parties decide to write will be perfectly enforced by a court, provided, of course, that it does not contravene any existing laws" (2005, p. 3).
[9] For other good examples see Dixit 2004; Milgrom, North, and Weingast 1990; and Greif 1989 and 1997.

before and had no relationship with anyone in the foreign city. If the Genoese merchant was cheated, he could take his case to the Hamburg guild court. That court had a strong incentive to treat the Genoese merchant fairly, because if the Genoese court found that its counterpart in Hamburg had failed to enforce the rules, it would expropriate the property of all of the Hamburg merchants doing business in Genoa, effectively ending trade between the two cities.

The creditability of this system depended on the magnitude of the rents created by long distance trade. In the first place, the value of the commerce between the two cities was what gave the guilds an incentive to protect each other's members. In the second, the profits that Genoese (or Hamburg) merchants received from their trading activities outside their own city made adherence to the guild's rules easier to sustain, because members had a lot to lose if they violated a prohibition. But that in turn meant that the creditability of the system also depended on each guild's ability to limit access to the trade to its own members. When the guild status of merchants became difficult to establish (for example, because growing numbers of traders provided cover for imposters), the system began to fall apart. The important point to underscore, however, is that it was the rents from the ongoing cooperation between two (or more) separate organizations that made private-order, third-party enforcement work.

There are many examples throughout history of how this kind of cooperation among organizations could be used for enforcement purposes. In the United States in the late nineteenth century, as we have already seen, cartels were not enforceable at law, but they were not yet illegal. Business people responded to this situation by devising private mechanisms to enforce cartel discipline. For instance, the railroads attempted to organize Standard Oil and several other important petroleum refiners into the South Improvement Company in 1871, an organization tasked with policing their agreement to maintain freight rates. In exchange for this service, the

refiners would receive a share of the resulting monopoly rents (Granitz and Klein 1996). Although oil producers' vehement (and potentially violent) protests prevented the South Improvement scheme from going into operation, similar arrangements effectively stabilized cartels in other industries. Margaret Levenstein has shown, for example, that bromine producers were able to use national distributors of fine chemicals to stabilize their collusive arrangements (Levenstein 1995). To give a very different kind of example, large landowners in nineteenth-century Sicily turned in the absence of effective government to an extralegal private organization, the Mafia, to protect their property rights (Gambetta 1993, Bandiera 2003, Dixit 2004 and 2009).

Douglass North, John Wallis, and Barry Weingast (NWW 2009) have extended the logic underpinning these examples to show how the rents generated by the cooperation of two or more organizations can limit violence. Think of leaders of two groups that control labor, land, capital, and other resources. The leaders agree to recognize each other's "rights" to control their respective groups and resources. Because the rents they receive from their groups will fall if violence breaks out between them, there is a range of circumstances in which each leader can credibly believe the other will not fight. Following the logic of organizations laid out earlier, the relationship between the leaders creates rents from non-violence that provide incentives for them to continue to cooperate. NWW call this organization among the leaders the "dominant coalition."

This idea is represented graphically in Figure 1. A and B are the leaders. The vertical ellipses represent the arrangements the leaders have with members their own groups (the a's and b's) and the labor, land, capital, and resources they control. The horizontal ellipse represents the arrangement between the leaders that constitutes the dominant coalition. This horizontal

arrangement is an adherent organization that is made credible by the vertical arrangements—by the rents the specialists receive from controlling their client organizations. But there is also a reciprocal effect. The existence of the agreement between members of the coalition enables each of them better to structure their client organizations, because they can call on each other for external support. That is, if the horizontal relationship between coalition members is credible, they can provide each other with the third-party enforcement that enables the vertical relationships they have with their clients to evolve into contractual organizations. The rules may be as simple as B always helps A, but the willingness of B to come to A's aid creates a defensible property right for A in the resources of A's group. Anthropologist who study the emergence of more complex societies out of simple hunter-gatherer societies find that these types of rules and arrangements defining the identity and privileges of elites are very common.[10]

Figure 1

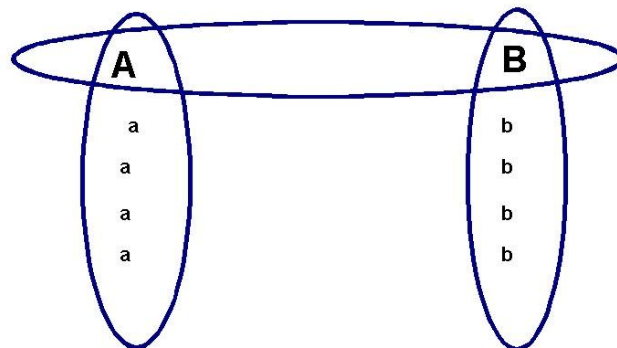

Figure 1 is simple in the extreme, but it enables us to surmount the logical problems that plague the literature on "violence specialists." Social scientists who consider the development of

[10] Johnson and Earle (2000), Earle (1997, 2003).

third-party enforcement typically postulate the existence of a powerful individual with a comparative advantage in violence. They proceed to define the interests of this individual and then theorize about the conditions under which he will honor his commitments to provide enforcement and protection.[11] The problem, however, is that the kind of violence that the third-party enforcer needs to wield is not individual violence but organized violence. A violence specialist cannot organize other violence specialists simply by threatening to beat them up or kill them, because a coalition of any two or more violence specialists can always defeat a single violence specialist, no matter how strong he is. Organizations that use violence must be held together by something other than coercion, and that something is rents.

The rents that hold dominant coalitions together are not restricted to rents from territorial political organizations; they are generated by other types of organizations as well. In Mexico under the dictatorship of Porfirio Díaz (1876-1910), for example, members of the elite who supported the government obtained special privileges that enabled them to profit from organizing industrial ventures. At the heart of this system were restrictions on entry into the banking sector that gave those who were politically well connected access to credit that other entrepreneurs lacked. The result was that industries that were competitively structured in other societies, such as cotton textiles, were dominated in Mexico by a relatively small number of large firms. The government then sheltered these oligopolies from foreign competition by enacting protective tariffs, further increasing the rents that coalition members could earn (Razo 2008; Haber, Razo, and Maurer 2003; Haber 1989).

---

[11] North's neo-classical theory of the state (1981 and 1990), Olson's stationary bandit (1993), Bates's fable of violence (2001, 2008), Bates, Greif, and Singh's organized violence (2002), Tilly's bargain between capital and coercion (1993), Greif's analysis of the podesta (2006), and Barzel's theory of the state (2002) all begin with the premise that an already existing organization of residents bargains with a single-actor violence specialist to provide protection and justice. Myerson (2008) captures the preoccupation of this literature in his title "The Autocrat's Credibility Problem and the Foundations of the Constitutional State."

During the early modern period, Western European governments similarly pursued a strategy of restricting access to profitable commercial opportunities, granting key supporters privileges to form special trading corporations, such as the British East India Company, and awarding those corporations monopolies over specific trading routes (Carruthers 1993).[12] Early modern governments also limited access to the corporate form more generally. Because the corporation's particular features (such as legal personhood, perpetual life, and limited liability) facilitated the pooling of capital for ends that promised high returns but required substantial investments, they were in and of themselves valuable privileges. These features depended on government-enforced rules, so in most cases governments controlled access to the form simply by declining to enforce the rules for companies they had not sanctioned. But in other cases, such as in England after the passage of the Bubble Act in 1720, governments made the formation of companies without approval a crime that at least potentially carried severe penalties (Harris 2000).

Although the rents that derived from privileged access to organizations helped to cement the dominant coalition, under some circumstances they could also be a source of instability. In the case of corporations, conflict could erupt among stockholders with different visions of the course the enterprise should follow, and when those stockholders were privileged members of the coalition, the disputes could have broader political implications. For example, mounting financial problems in the British East India Company during the early seventeenth century spurred angry shareholders to challenge the existing leadership. This conflict took on a political dimension when the disgruntled faction enlisted the support of King Charles I, who after being rebuffed by the company's leaders, pursued policies that undermined the company's privileges

[12]For histories of the various trading companies chartered during this period, see Chaudhuri 1965; Phillips 1990; Gepken-Jager, van Solinge and Timmerman 2005.

(Chaudhuri 1965; Scott 1910). Saumitra Jha (2010) has shown that shareholders in the British East India and other trading companies provided key support in Parliament for votes that precipitated the English Civil War (see also Brenner 1993).

More generally, the returns earned by members of the dominant coalition through their privileged access to organizational rents can themselves be an incentive to rebellion, encouraging factions within the dominant coalition, or even excluded groups, to stage coups if they think the rents to be gained by wresting control of the government exceed the costs of mounting a successful revolt. Even states that seem strong in terms of the force they are routinely able to mobilize can be highly dependent for their persistence on the members' perceptions of the value of remaining in the dominant coalition. The government of Hosni Mobarak in Egypt provides a dramatic and recent example of how quickly apparent strength can evaporate if coalition members' interests shift. The Mubarak regime was autocratic and, by most measures, extremely powerful. And yet when the Army was unwilling, in the face of popular protests, to follow Mubarak's command to shoot people in Tahrir square, the government suddenly weakened. There was no change in the internal capacity of the Mubarak government, but there was a dramatic change in the configuration of the dominant coalition in Egypt. In that coalition the Army was only subordinate to the government if it agreed to be; when it declined to follow Mubarak's lead, the regime collapsed.[13]

Although it is too soon to see whether the "Arab spring" will dramatically transform Egyptian society, historically such upheavals have rarely affected the basic structures of power.

---

[13] Currently, the Army in Egypt still appears to be strong, but this strength in turn depends on the internal organization and cohesion of the groups that make up the Army. It is tempting to think of the current Egyptian Army as a "single actor" with a well-defined objective function, but the military council in Egypt is clearly itself a coalition. The Egyptian case is a very clear example of why it is the internal dynamics of the dominant coalition, rather than some notion of government capacity, matter so much to the ordering of society.

The rent-creating organizational arrangements that support dominant coalitions are equilibriums in the sense that, whenever such arrangements are destroyed, similar ones generally emerge to take their place. The collapse of one dominant coalition thus leads to the creation of another that functions in much the same manner, restricting access to organizational rents for the benefit of the new ruling group. When Díaz's regime fell during the Mexican Civil War, industrial production did not suffer because the rebels kept the factories in operation for their own purposes (Haber, Razo, and Maurer 2003). Latin American history is replete with examples of one "caudillo" following another in a seemingly endless cycle of corruption, violence, and more corruption (North, Summerhill, and Weingast 2000). The plight of African nations since their independence from colonial rule has been sadly similar. Robert Bates's work on Africa (1989, 2001, and 2008) shows a clear pattern that parallels the Latin American experience.

1.4. States, Governments, and Organizations

The term "state" is used in many different ways in both ordinary conversation and scholarly discourse. Adopting one common meaning, we define the state to be the combination of powerful interests and organizations that orders relationships throughout a society—in other words, essentially the dominant coalition depicted in Figure 1. Although this definition may seem commonplace, it is important to emphasize how different our view is from many others in the literature. Weber (1948) classically defined the modern state as the organization with a monopoly on the use of legitimate violence, Tilly (1990) as the organization with a priority on the use of violence, and North (1981) as the organization with a comparative advantage in violence. In our analysis, neither A nor B can be said to have an advantage in the ability to

threaten violence in the way described by these theories. Rather, as we have shown, credible third-party enforcement of rules is sustained by the interaction of their organized interests.

To the extent that leaders have a comparative advantage in violence vis-a-vis their clients, this advantage arises because the rents from their relationship create credible incentives for each to coordinate with the other—not because the leaders are exogenously endowed with a superior ability to fight. The rents from cooperation are themselves dependent on the rents each leader earns from controlling his group. So, once again, we see that dominant coalitions are adherent organizations of organizations held together by the rents members earn from belonging to them. The third-party enforcement they provide depends critically on the existence of at least two organizations with interlocking, rent-generating interests.

States (that is, dominant coalitions) began, historically, as the highly personal creations of powerful individuals. As the rents from cooperation grew, however, states became more complex, and the methods of regulating relationships and adjudicating disputes among coalition members became more formal. Here, as Dixit (2009) reminds us, we must be conscious of the difference between government and governance. Governance within the dominant coalition does not necessarily require a government. We define a government to be one (or more) of the many governance organizations that may exist within the dominant coalition. What distinguishes governments from these other organizations is their publicness.

Who and what constitutes the dominant coalition (the state) in many, perhaps most, societies is a difficult empirical question to answer because many powerful actors are not visible. Governments, by contrast, are highly visible organizations, and it is precisely because of their publicness that they are good at facilitating coordination. The presumption in the literature on states that we cited above is that a government’s ability to coordinate depends on its ability to

coerce—that is, on the state's monopoly on violence. If a government primarily coordinates through the provision of rules, and if the credibility of rules depends on the government's ability to enforce adherence, so the logic runs, then the government's coordinating function depends on its coercive ability.

Certainly governments sometimes enforce rules through the threat of violence; there is little doubt that coercion is an important aspect of many governments. But, we would argue, within the dominant coalition the government's ability to create and enforce rules depends as much or more on the value of the coordination that the rules bring about as it does on the extent to which the government can discipline coalition members. Indeed, in societies where non-government organizations within the dominant coalition are particularly strong relative to government organizations, the government's role may be primarily to articulate focal points rather than enforce agreements.

The creation and enforcement of rules cannot be a monopoly of governments. Organizations always have the ability to enforce some rules for each other.[14] The particular mix of public and private ordering that occurs in any society will depend on the relative value of these alternatives to its members. Nonetheless, we hypothesize that the extent of public ordering is systematically greater in open- compared to limited-access societies. Where access to organizations is limited, many of the rules that govern relationships within the dominant coalition are neither publicly stated nor publicly enforced. After all, in these societies the dominant coalition relies on its ability to discriminate among individuals and the organizations to

---

[14] This does not mean that organizations can enforce any rules they want. The rules must be consistent with the interests of the organizations involved in the private ordering arrangements.

which they belong. In open-access societies, by contrast, the same rules apply to everyone, and so the public ordering function of the government must expand.[15]

Limited-access societies tend to be unstable. The private nature of many of the rules that govern important relationships means they can shift quickly and dramatically along with the interests of groups in the ruling coalition, as the recent experience of Egypt dramatically illustrates. Weber, North, and Tilly have explained the greater stability that generally comes with the transition to open access in terms of the government's acquisition of a monopoly of the means of violence. It is never clear, however, why powerful groups in a society would be willing to concede this monopoly to government. In most societies, such a move would be potentially suicidal, for any organization could fall out of the part of the dominant coalition that includes the government. Moreover, though it is certainly the case that the government's coercive power increases in open-access societies, both in absolute terms and relative to private organizations, that cannot be the whole story. Simply expanding the coercive power of government is not, on its own or in historical terms, a particularly good outcome, since a powerful government organization is typically a more repressive and exploitive one.

How we understand the mix of public and private ordering, and the dynamic relationship between the two, is the central problem of this conference. The conventional view conceives of the modernization process as driven by the progressive growth in the capacity of government to provide credible institutions of public ordering. Public ordering is viewed as a superior substitute for private ordering that facilitates higher levels of social development and economic growth. Of course, there is no doubt that the government's role in public ordering increased dramatically in the nineteenth century. But it is also important to recognize that the number and

---

[15] We do not literally mean everyone; there is still discrimination in most open-access societies. The question of how open open-access societies have to be for the new equilibrium to be sustainable is a question for further research.

scope of private organizations and private-ordered relationships also increased dramatically at the same time. We want to explore the idea that these two trends were related—that in the early nineteenth century an expanding government came to function as a complement to private ordering. As a result of this complementarity, we would like to suggest, the value of public ordering to elite organizations suddenly increased, so that elites were willing to support a wider provision by government of the requisite organizational tools—first to themselves, but willy-nilly eventually to all citizens. Britain and the United States were the first countries to undergo this radical transformation to open access, and in the historical sections that follow we examine their experience.

2. Transitions from Limited to Open Access

Douglass North is fond of saying that it took four hundred years of institutional change to produce today's rich democratic societies, and there is no question that in an important sense this statement is correct. One can easily lay out a long series of historical events in Europe, and especially in England and its colonies, that gradually paved the way for the emergence of sustained economic growth and civil liberties in the nineteenth century.[16] At other times in his career, however, North has taken a very different view and attributed a primary causal role to the sudden, cataclysmic changes associated with the Age of Revolutions. In the case of Britain, for example, he and Weingast (1989) have argued that the upheavals that began with the English Civil War and culminated in the Glorious Revolution ushered in a new era of limited government, providing the enhanced security for property rights needed to stimulate private

[16] Herbert Butterfield's famous *The Whig Interpretation of History* (1931) was a reaction against such

investment.[17] Weingast has attributed a similar importance to the American Revolution and the writing of the Constitution of the United States.[18]

Whatever truth there is in these two perspectives on historical change, if our purpose is to understand the transition to open access, both the gradualist view and the focus on revolutionary transformation miss important pieces of the puzzle (as North, Wallis, and Weingast recognized in their 2009 study). In the next section we show that neither the Glorious Revolution in Britain nor the ratification of the Constitution in the United States ushered in an open access social order. To the contrary, the regimes that followed these events relied on limited access to organizations to stabilize their positions in much the same way as had those that preceded them. In the following section we then examine the transition to open access in both countries. We focus in particular on the different experiences of the various U.S. states in order to underscore how difficult it is to view the transition as the necessary and inevitable consequence of the long sequence of institutional changes that came before.

2.1. The Age of Revolutions

The Glorious Revolution in Britain may have shifted the balance of power between the King to Parliament, but its consequences for the organizational dynamics with which we are concerned was much less profound. The coalitions struggling to control Parliament in the aftermath of the rebellion sought to reinforce critical alliances in the time honored way—by granting favored groups monopoly rights and other lucrative privileges. Recent work on the Parliamentary acts creating organizations to build canals and turnpikes during this period

[17] Although other scholars have differed in their views of the precise timing and mechanism of the change, there is general agreement that the Glorious Revolution ushered in a new period of limited government in Britain. See Brewer 1990, Stasavage 2003, and Pincus 2009.

[18] See, for example, his essay with Sonia Mittal and Jack N.. Rakove in Irwin and Sylla 2011.

suggests that access to these privileges remained limited to those connected to the party in power (Bogart and Oandasan 2012). But perhaps the best indicator of how little had changed was the Bank of England. The bank had its origins in the dire need of the new government for funds. Parliament had secured a long-term loan of £1,000,000 in 1693 but within a year had burned through the whole amount. Lenders were understandably reluctant to commit more money to the government, but Parliament succeeded in attracting an additional £1,200,000 in loans by agreeing to incorporate the subscribers as the Bank of England (North and Weingast 1989). The government's debt, which was secured by additional taxes, served as the bank's capital. On the basis of that capital the bank could issue notes, make loans, and accept deposits, with the subscribers protected from excessive losses by limited liability. Crucially, Parliament limited the bank's charter to a term of eleven years. To the extent that the bank proved valuable to the subscribers, this short lifespan gave Parliament additional opportunities to extract loans in exchange for privileges. In fact, Parliament needed funds again before the eleven years were up and was able to secure in excess of $1,000,000 in new loans in exchange for an exemption for the subscribers from taxes on their bank stock and a promise not to set up any competing banks. In 1708, in exchange for still more loans, Parliament granted the Bank of England an effective monopoly by prohibiting any joint stock company with more than six partners from doing business as a bank. Over the next century Parliament repeatedly renewed the bank's charter, confirming its privileges in exchange for loans. It also negotiated similar arrangements with the New East India Company in 1698 and the South Sea Company in 1711 (Broz and Grossman 2004).

The American Revolution was to a large extent a reaction against this kind of favoritism, which tended to work to the disadvantage of all but the most well-connected colonists. Like the

Glorious Revolution before it, the American Revolution has traditionally been viewed as an institutional game changer. For the achievement of open access, however, it is important not to overstate the extent of the transformation. Although the colonists had a well worked out understanding of the ways in which privileged access to organizational rents could undermine liberty, they had no positive theory of how a government that did not limit such access might work. The only idea they had about how to prevent their society from heading down the primrose path to tyranny was to limit the size and power of government. After the first government they created under the Articles of Confederation proved unworkably weak, they fumbled their way toward a stronger alternative with three branches that would check and balance each other to prevent abuses of power (Wood 1969). There was little in this setup to secure the allegiance of the rich and powerful, however, and so federalists like Alexander Hamilton consciously resorted to the old methods, offering privileged access to organizational rents to wealth holders to link their interests to the success of the new government (Wallis 2008). Hamilton's use of the national debt to commit the wealthy to support "the fiscal arrangements of the government" is well known.[19] His proposal for the formation of a national bank, modeled on the Bank of England, had similar aims. Hamilton promised potential investors that the bank would have a monopoly in the sense that "no similar institution shall be established by any future act of the United States," so long as the bank continued in existence.[20] Shares of the bank would be payable in the public debt of the United States, unifying the interests of the debt holders and the bank's subscribers. In addition, Hamilton imitated the Bank of England's symbiotic relationship with Parliament by proposing that the Bank of the United State be chartered for only

---

[19]Hamilton, "On Public Credit," 11.

[20] Hamilton, "On the Establishment of a National Bank," 124. The Continental Congress had chartered the Bank of North America in 1781, and Hamilton devoted more than a third of his text to explaining why it is was not a violation of the monopoly principle for Congress to charter a new bank. In his view, the state of Pennsylvania had crippled the Bank of North America by unilaterally revising its charter. The states, of course, could also charter banks, but in the early years they too adhered to the monopoly principle. See below.

a limited term. Because the bank "must depend for its renovation, from time to time, on the pleasure of the government, it will not be likely to feel a disposition to render itself, by its conduct, unworthy of public patronage." At the same time, the government, "in the administration of its finances, has it in its power to reciprocate benefits to the bank, of not less importance than those which the bank affords to the government …"[21]

From the beginning there was significant opposition to Hamilton's plans, but the only alternative that critics had to offer was the eradication of organizations like the national bank that they viewed as sources of corruption (Wallis 2008).[22] Rather than increase the number of organizations, their solution, in other words, was to limit them even further. Although Congress enacted Hamilton's proposal over the misgivings of the Jeffersonians in Congress, when the bank's charter expired in 1811, the latter were in power and they refused to renew it (Hammond 1957). The Jeffersonians also managed completely to scuttle another of Hamilton's plans—to offer prizes and bounties to support manufacturing endeavors (Elkins and McKitrick 1993). The one glimmer of an alternative way of organizing society was the creation of the patent system which rewarded inventors who devised novel and useful technologies with temporary monopoly rights. The first Congress deliberately expanded access to this protection for intellectual property by dramatically lowering the cost of patents relative to Britain. This change, however, was more a reflection of the founders' rejection of the British idea that only members of the elite were likely to come up with valuable new technological ideas than it was a commitment to opening access to organizational rents (Khan and Sokoloff 1998; Khan 2005).

---

[21] Hamilton, "Report on a National Bank," 119.

[22] To be fair, Hamilton had been a revolutionary too. He understood the problem of corruption, and his solution was to build some checks and balances, such as rotation of directors and a graduated voting scheme, into the bank's governance structure.

Under the Jeffersonians the federal government remained small, confining its activities for the most part to delivering the mail, granting patents, collecting customs duties, and maintaining a small army and navy. Indeed, it did so little during this period that no one wanted to serve in it. From one third to one half of the Congress left every two years, mostly as a result of resignations. The departing members were not retiring from politics, however. Most went on to hold office at the state or local level, where there was a lot more going on (Young 1966). And what was going on was what governments had done since time immemorial: creating organizations that generated rents for the favored few who were allowed to participate in them. Massachusetts, for example, chartered a bank in 1784 that bore the commonwealth's name, served as the fiscal agent of the government, and even issued notes bearing the commonwealth's seal. New York, Rhode Island, and several other states followed suit over the next decade, chartering banks with effective monopolies in their jurisdictions (Handlin and Handlin 1969, Gras 1937, Bodenhorn 2003, Lamoreaux 1994). Many states also granted monopoly privileges to groups of investors who promised to undertake badly needed transportation projects (Hartz 1948, Handlin and Handlin 1969, Majewski 2006). The merchants who contracted to build a bridge over the Charles River at Charlestown, for example, received a charter from the Massachusetts legislature giving them the right to collect tolls over this route for forty years (Kutler 1990). New York and other states granted Robert Fulton and Robert R. Livingston a monopoly on navigation by steamboat in their waters (Cox 2009). Some states also conferred extensive privileges on favored manufacturing ventures. New Jersey, for example, granted the Society for Useful Manufactures (SUM), a textile company formed by (among others) Alexander Hamilton, boons in its charter that included the right to raise funds through a public lottery and exemption for the company's employees from taxes and military service (Maier 1993).

1.2. Opening Access

Whatever changes the upheavals of the Age of Revolutions may have brought about in Britain and the United States, in the short run they had relatively little effect on the government's willingness to give up control over the formation of organizations. By the mid-nineteenth century, however, both countries had shifted dramatically toward open access. The change occurred in very differently in the two countries. In Britain, about one hundred and fifty years after the Glorious Revolution, the shift occurred suddenly and with remarkably little conflict. Britain had relatively few corporations in the early nineteenth century because it was it was easy for coalitions of interested parties to block charters in Parliament. Most large-scale businesses, as a result, took the form of unincorporated joint-stock companies. These were essentially partnerships that had developed contractual practices that allowed them to concentrate managerial authority and function, for the most part, as if they were legal persons. They were, however, a second-best solution to the problem of pooling capital, and during the mid 1820s the courts struck them a series of blows by handing down decisions that interpreted the Bubble Act in ways that cast doubt on their legality. Worried entrepreneurs responded by deluging Parliament with petitions for corporate charters, and, overwhelmed by the sheer number of petitions, Parliament repealed the Bubble Act in 1825 (Harris 2000; Freeman, Pearson, and Taylor 2012).

Not surprisingly, the number of joint-stock companies surged over the next couple of decades, but so did the number of companies that went bust. Entrepreneurs who wished to form new companies, as well as wealth holders with assets to invest, wanted something done about the situation, and consensus built within the elite for legislation that would regulate the hodge-podge of private companies soliciting investments from the public without killing them off. The

composition of Parliament had already shifted in favor of business interests in the wake of the Reform Act of 1832. In 1844, responding to the growing demand for action, Parliament passed an act granting corporate status, though not limited liability, to any company that registered, met certain minimal requirements, and promised to file regular financial statements. It completed the transition to general incorporation by adding limited liability in 1855 and 1856 (Harris 2000; Taylor 2006; Freeman, Pearson, and Taylor 2012).

Around the same time, Parliament removed the barriers to competition that had protected the great monied corporations of the previous century. Lobbying by eager would-be competitors secured the repeal of the East India Company's monopoly in 1833 (Fichter 2010). Parliament similarly removed the limits on banking partnerships built into the Bank of England's charter, first for institutions outside London in 1826 and then everywhere in 1833. It completed the shift to openness in banking with the passage of a general incorporation law without limited liability for banks in 1844 and with limited liability in 1855, and then, over the next several years eliminated the remaining special restrictions on banks by putting them under the domain of the general company law (Broz and Grossman 2004, Grossman 2010).

The period also witnessed a shift in Britain toward greater toleration of organizations of all kinds. Although men and women had been forming growing numbers of clubs and other kinds of voluntary associations since at least the mid-seventeenth century, organizations that posed a challenge to the established political or religious order had been forced to lead a shadowy existence on the margins of legality, protecting themselves from repression by touting their charitable activities or hiding their real purpose under a cloak of secrecy or frivolity (Lund 2002, Clark 2000, Shields 1994, Jacob 1981, Albers 1993). During the era of the French Revolution and the Napoleonic Wars, there was an increase in outright repression, but by the

middle of the nineteenth century it had largely abated (Emsley 1985, Thompson 1964, Hone 1982, Seed 1985, Rimlinger 1977; Gregory 2000). Although the Church of England continued to dominate religious life, dissenting congregations organized freely in opposition to its establishment, and though they did not succeed, were able to free themselves of most of the legal disabilities that had hampered their activities in the previous century. The repeal in 1824 of Combination Acts (passed in 1799 and 1800, they had tarred labor organizations as criminal conspiracies) delivered unions from the taint of illegality, though it was not until the 1870s that they began to get access to standard organizational tools. In other walks of life, clubs and voluntary associations of all kinds—charitable, educational, recreational, political—mushroomed in the early nineteenth century and then continued to spread as the century progressed (Gunn 2000, Morris 1990, Davidoff and Hall 1987, Rimlinger 1977).

In the United States, the move to greater openness followed a much different path than it did in Britain. Indeed, because the key institutional changes occurred at the level of the states rather than the federal government, the move to open access in the U.S. followed multiple paths. Unlike Britain, moreover, where the shift occurred around the same time for most types of organizations, in the U.S. the change could be much more uneven. Thus Pennsylvania passed a general incorporation law for churches as early as 1791 but did not pass a similar act for banking until 1860 (Hartz 1948, Bloch and Lamoreaux forthcoming). New York passed a temporary general incorporation law for manufacturing in 1811 to stimulate local industry during the period when trade with Britain and France was embargoed. Scheduled to expire in 1816, the legislature hesitated to renew it, passing a couple of one-year extensions before letting the act lapse only to revive it permanently in 1821 (Seavoy 1982, Lamoreaux and Harris 2010). New Jersey passed what was almost a verbatim copy of the New York manufacturing statute in 1816 but then

repealed it in 1819. Not until 1846 did New Jersey make the corporate form generally available for manufacturing purposes (Cadman 1949; Harris and Lamoreaux 2010). As a general rule, Southern states lagged behind their Northern counterparts in passing general incorporate laws, with Georgia and Texas first making the form freely available for manufacturing firms only in the 1870s (Hamill 1999).

Because the experience in the U.S. was so varied across both states and types of organizations, in the discussion that follows we focus on just one sector of the economy: banking. We chose banking because of the key role that the development of the financial system played in economic growth. We also chose banking because control of the rents generated by financial organizations was typically viewed by the coalition in power as critical for its success and, as a consequence, the shift to open access in banks was particularly fraught politically. In no state before the 1830s was there open access to banking. Several states, including Indiana, Missouri, and Kentucky chartered state owned single monopoly banks with branches in different parts of their territory. Most, however, issued charters to private investors in different locations, with each town or city (except the very largest) getting only one institution.

In Massachusetts, for example, the Federalists controlled the government for most of the period from the Revolution to the late 1810s. Massachusetts chartered 23 banks between 1784 and 1810 (see Figure 2), and all the charters went to Federalists. As in the case of the Bank of England and the Bank of the United States, the charters were of finite duration. All were due to expire in 1812. Members of the rival political coalition, the Republicans, had pressed for additional charters that would benefit members of their party, but the Federalists blocked their efforts. The tables were turned, however, when the Republicans gained control of the state government in 1810 and 1811. The Republicans were able to push through two bank charters—

one for a bank in Salem, the other for a massive new State Bank with a capital of $3,000,000, by far the largest in the commonwealth. The government had the option of investing an additional $1,500,000, in which case it would be able to appoint a third of the bank's directors. It never made the investment, but another provision proved to be of great importance: the government was to receive a one-percent annual tax on the bank's capital (Lu 2012, Handlin and Handlin 1969, Hammond 1957, Goodman 1964).

Although Federalists managed to secure a few shares in the institution, the bank distributed its stock "as extensively as possible among the Friends of Government" (Stetson 1891; Knox 1908). The Republicans trumpeted the chartering of the State Bank as a reform measure that created a public institution to replace the private banking privileges that the Federalists had allotted to their favorites, and they followed up by refusing to renew the charters of the existing Federalist banks. Now it was Federalists' turn to denounce the Republicans' monopoly of banking.[23] To make matters worse, in 1811, the Republican governor, Elbridge Gerry, led a vigorous attempt to entrench the Republicans in power, including the famous effort to redraw the state's senatorial districts that inspired opponents to coin the term "Gerrymandering" to mean redistricting for partisan advantage (Lu 2012; Griffth 1907, 17-21, Austin 1828, 322; Dean 1892, 374-83; Hofstadler 1968, 250).

---

[23] *Boston Gazette*, June 24, 27, 1811; *Massachusetts Spy*, Aug. 14, Sept. 4, 11, 13, 24, 1811; *New England Palladium*, June 16, 21, 1811; Stetson (1891), 12-13. See also Lu (2012).

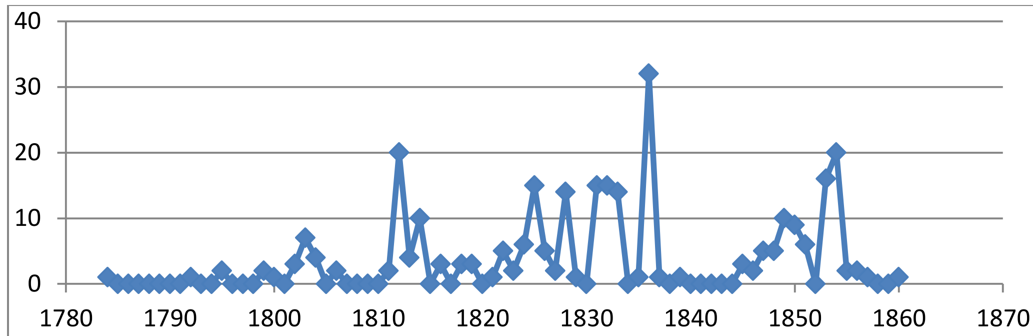

Figure 2. Bank Charters Granted in Massachusetts by Year

Source: Sylla and Wright (2012).

The Federalists nonetheless regained control of the executive and legislative branches in 1812. Although they moved immediately to grant new charters to their old banks (see Figure 2), this time they brokered an arrangement with the promoters of the State Bank whereby the latter would continue to pay the one-percent tax to the state on its capital on condition that all the other banks in the state did likewise. In fact, the deal went even further and gave all the banks charters that (except in the details of their location and capital stock) were identical to that of the State Bank.[24] The tax on bank capital was such a tremendous boon to the commonwealth's finances that opposition to additional banks largely evaporated. In 1830, the first year for which data is available, this one tax on banks amounted to fully 61 percent of the state's revenue. Massachusetts did not have to impose *any* property or poll taxes on its citizens for half the years between 1826 and 1855 (Wallis, Sylla, and Legler 1994, 126; Sylla, Legler, and Wallis 1987, 394).

Few new banks were chartered in Massachusetts during the turbulent war and depression years that followed the creation of the State Bank, but as economic conditions improved during the early 1820s the demand for charters increased and the legislature responded by granting

[24] See *Laws of the Commonwealth of Massachusetts from February 28, 1807, to February 27, 1813* (Boston: Thomas and Andrews, 1813).

them. This is not to say that bank charters were never again a hot-button political issue, just that they were no longer reserved to support the elite coalition. In some years, as can be seen in Figure 2, the legislature refused to grant any requests for charters, but the logjam usually burst the next year with a surge of approvals. The formal shift to open access was almost a bureaucratic afterthought. As the number of charters increased, so did burden on the legislature, which moved to streamline the process—first in 1829 by enacting a law that establish a template for charters, and then in 1851 by passing a general incorporation law for banks. In Massachusetts and most of the other New England States, the process of granting corporate charts had been so effectively depoliticized by that point that there was no constitutional mandate for general incorporation. As a consequence, legislatures in the region continued to grant charters by special act long after the passage of general laws had otherwise routinized the process (Hamill 1999, Evans 1948, Knox 1908).

It is difficult to escape the suspicion that the fortuitous imposition of a tax on bank stock in 1812 changed the political calculus in ways that facilitated the shift to open access. Legislators were faced with the alternative of chartering additional banks or imposing onerous taxes on their constituents, and not surprisingly they found the former choice more attractive. In New York, where there was no similar tax on bank capital, the path to open access was very different, though the early history of banking was much the same. In the years following the Revolution, the Federalists were in power, as they were in Massachusetts, and they allocated bank charters only to members of their coalition. As in Massachusetts, moreover, when the Republicans gained power, they similarly restricted charters to their own followers. In 1812, the Republicans were still dominant in New York, and Federalists seeking a charter for a huge new bank in New York City (the proposed capital was $6 million) hired some Republican politicians

to spread money around to further their cause. In the ensuing bribery scandal, the governor, who opposed the charter, suspended the legislature on the grounds that "the confidence of the people in the purity and independence of legislation [would] be fatally impaired" if the legislation were to pass.[25] The prorogue of the legislature did little good, however. The assembly continued to dole out charters to political favorites, and rumors of bribery continued to spread (Bodenhorn 2006).

Reformers as yet had no vision of the possibilities of an open-access order, and they responded to this environment by inserting into the new state constitution of 1821 a provision requiring a two-thirds vote of both houses of the legislature to charter a bank. By making charters more difficult rather than easier to obtain, however, they provided politicians with an even more useful device to reward supporters and punish opponents—a device that Martin Van Buren used very effectively to build his political faction (which in the 1830s became the core of the Democratic Party in New York) into a powerful machine. During the economic boom of 1830s the general assembly received on average about 70 petitions for banks a year, but under the machine's tight control only about ten percent of that number ultimately received charters. Not surprising, the large number of rejections helped fuel opposition that was brewing to Van Buren's "Albany Regency." When the collapse of the banking system in the Panic of 1837 finally brought the machine down, the opposition (now called the Whig Party) met the pent-up demand for charters and, at the same time, insured that the Democrats would never again be able to use bank charters for political purposes by passing New York's famous free banking law in 1838. Now anyone who met the law's basic regulatory requirements could form a bank by a simple registration process (Benson 1961, Hofstadter 1969, Bodenhorn 2006, Knox 1908).

---

[25] Message of Governor Daniel D. Tompkins to the New York General Assembly, 27 March 1812, in *Messages from the Governors*, ed. Charles Z. Lincoln (Albany, NY: J. B. Lyon Co., 1909), Vol. 2, 711.

Other states responded to the financial crisis of 1837 differently. For example, in Pennsylvania, whose early banking history was much the same as that of New York and Massachusetts, reformers responded as they had in New York in 1821—by making it more difficult to secure charters. The new state constitution passed in 1838 actually forestalled open access by stipulating that "no law hereafter enacted shall create, renew or extend the charter of more than one corporation" (Knox 1908, 450-51). Tennessee's response was to create a huge state-run bank, the Bank of Tennessee, which not surprisingly was controlled by politicians (Knox 1908, 653). Texas's first state constitution in 1845 prohibited the legislature from chartering banks altogether, declaring that "no corporate body shall hereafter be created, renewed or extended, with banking or discounting privileges" (Knox 1908, 620).

More generally, the response to crisis was to restrict access to charters rather than to open the process up (Knox 1908). Open access did not emerge as a result of a process of democratic reform. The political activists of the time viewed banks as instruments of a monied elite and monsters of privilege, and they opposed chartering them, period. Rather, whenever and wherever open access emerged, it was the product of an elite bargain. Entrepreneurs who wished to garner banking privileges found they had to win control of the political process to do so, and in a few key states they slowly came to realize that their interests would be better served if people of their ilk on both sides of the political divide were able to secure bank charters, regardless of whether they occupied positions of political power. In Massachusetts, this realization came early, aided by the beneficial effect on state finances of the tax on bank capital; in New York it came considerably later, awaiting the fall of the Albany Regency after the Panic of 1837. Whether it would have come in all the other states independently is not clear. New York's free banking law had an important demonstration effect, however, and seventeen other

states adopted similar statutes in the years that followed, most of them during the economic expansion of the 1850s (Ng 1988). As for the remaining states, open access came via the federal government, and again not as a triumph of reform. Desperate to finance the government debt during the Civil War, Congress passed a general incorporation law for banks modeled on New York's free banking law, making it possible for entrepreneurs to form banks in any state of the union regardless of state law (James 1978, White 1982).

If Americans were going to use the allocation economic privileges to secure political coalitions, then those economic privileges would forever be held hostage to the party in power. American elites came to see that the economic value of the organizations they led and belonged to would be increased if they changed the political dynamics around the formation of those organizations. If bank charters were granted or revoked based on who was in political power at a particular time, the economic value of the banks could only be secured by either making sure that one faction remained continuously in power (and limited access) or both parties agreed to allow the other to form and sustain banks (and opened access). Interestingly, Americans did not see this as an economic problem, but a political problem (Wallis, 2005 and 2006). If democratic elections decided which economic group triumphed, democracy would always be corrupted by political attempts to manipulate the economy.

Some Ideas for Research

The forgoing sections suggest three sets of research questions around which we would like to organize discussion at the conference. Here we will say a little about each in turn, providing some suggestions for further investigation that we hope will connect with your own

research concerns. Our suggestions are simply illustrative. They are not meant to be comprehensive, and they certainly should not be taken as directive.

How should we think about organizations, their internal structures, and the relationships among them? How do relationships among organizations affect their internal structure? How do organizations function differently in limited access environments than under open access?

Organizations can be structured hierarchically, or they can be structured in a relatively egalitarian manner. Presumably the choice is to some extent a matter of size; large corporations with thousands of small shareholders are more likely to be hierarchical than small closely held corporations. But the internal structure of organizations may also be affected by the nature of the social order—that is, all other things being equal, the way in which organizations are structured (or more accurately, the distribution of organizational structures of various types) may be systematically different in limited-access compared to open-access societies. We think it is important to think formally about why this might be the case—how privately ordering might function differently in the shadow of the law than outside of law (Dixit 2004), how the purpose of organizations might change when they can be freely formed (the "garbage can theory" of Cohen, March and Olsen 1972; Gibbons 2003 and 2005), and how the leadership function might evolve (Ahlquist and Levi 2006 and 2011). We also think it is worth thinking about how such developments might matter for various economic, social, and political outcomes.

From the standpoint of the economy, there is already a growing body of empirical work suggesting that such differences are real and that they affect performance. For example, John Haltiwanger and several coauthors have produced a series of studies documenting differences in

the dynamics of firms over time and across industries and countries.[26] The papers are built around cross-country comparisons often incorporate cases from both sides of the limited/open access divide, and the differences they find are highly suggestive for our purposes. Thus Bartlesman, Haltiwanger, and Scarpetta (2011) examine the relationship between size and productivity, controlling for industry. As one might expect, large firms tend to be more productive than small firms, but the relationship between size and productivity was much stronger for the five developed countries in their study than for the three Eastern European transition economies (though in the latter cases the correlation tended to get stronger over time as the transition process proceeded). The authors take the relationship between size and productivity to be a measure of the allocative efficiency of the economy. In countries where large firms benefited from privileges granted to politically well-connected entrepreneurs, the relationship between size and productivity was comparatively weak. However, as the transition economies moved towards policy regimes that enabled greater entry, the economies became more competitive, and the relationship between size and productivity (that is, allocative efficiency) increased.

Another example comes from the work of Nicholas Bloom, Raffaella Sadun, and John Van Reenen on the impact of managerial structures on productivity.[27] Using a large, multi-country survey that includes both limited- and open-access societies, the authors examine the degree of "centralization" of a firm's management. They find that firms with centralized management tend to have relatively flat hierarchies and to be constrained in how big they can grow. By contrast, firms with decentralized management structures benefit from a more

[26] These studies include Haltiwanger (2012), Foster, Haltiwanger, and Syversen (2008), Foster, Crizan, and Haltiwanger (2006),and Bartlesman, Haltiwanger, and Scarpetta (2009, 2010, 2011).
[27] Papers include Bloom, Sadun, and Van Reenen (2007, 2009, 2010a, and 2010b), Bloom and Van Reenen (2006, 2007, and 2010), and Bloom, Shweiger, and Van Reenen (2011).

productive specialization and division of labor within the firm and are consequently able to grow much larger. To understand the patterns of centralization versus decentralization that they observe across countries, they focus on the role of trust ("if a CEO can trust his senior managers he will be more willing to decentralize decision making–for example, the threat of theft is lower") and the rule of law ("there is more chance of getting stolen money back (and also less chance of theft in the first place) if contracts are well enforced and respected") (Bloom, Sadun, and Van Reenen 2009, 2). From our perspective, their finding that management practices are indeed related to these variables is highly suggestive because both lack of trust and rule of law are indicators of weak third-party enforcement. In other words, Bloom, Sadun, and Van Reenen's results (along with those of Haltiwanger and his co-authors) provide intriguing evidence that management practices may be endogenous to the type of social order—that organizations function differently in a limited access environment than they do in countries where organizational tools are freely available.

One potential problem with such studies is that their empirical methodology—particularly the estimations with continuous variables—implies that the path to modernity is easier than it seems to have been historically. There may be a continuum of managerial practices, but that does not mean that movement along the continuum is incremental or, if it is, that such movement is sustainable. Our hypothesis that limited and open access are two social equilibriums suggests that gradual and continuous change may be problematic in the world, even if it is not in our econometrics. This possibility brings us to the second set of questions.

How do societies transition from limited to open access? How is it that elite organizations with strong interests in limiting support for organizations decide to adopt rules that apply equally to

all organizations? What role does the shift to the public provision of third-party enforcement (government) play in the transition process?

The work by Dixit and Greif discussed above suggests that private organizations can provide third-party enforcement to protect property rights and facilitate long-distance trade, but it also suggests that these mechanisms do not easily scale up—that their effectiveness is contingent on limiting access. More generally, when groups benefit from the private provision of organizational discipline, they have no incentive to push for the public provision of laws that govern organizations. As Weingast (2010) has argued, in most societies elites find it in their interest to support a political and economic system in which privileges are unique and personal. If the most powerful organizations in a society are able to secure their relationships through private ordering, it is not obvious that they will gain from opening up access to contractual tools to the wider society. In present-day Indian, for example, the so-called "bollygarchs" have built globally competitive businesses on a foundation of family networks and close and repeated interaction among a small number of very large firms. These firms use the government as a third-party enforcer relatively infrequently.[28] Do the bollygarchs provide third-party enforcement to each other because Indian society has weak institutions, or are Indian institutions weak because the bollygarchs are able to provide private third-party enforcement without recourse to government?

If elites have an interest in maintaining limited access, how do things change? The historical narratives in the preceding section suggest that the transition to open access occurred when competing groups of elites all had significant organizational interests to preserve and saw

[28] See the special issue on India in the *Economist Magazine* (22 Oct. 2011). Khanna and Palepu (2000) examine the role of family based business groups in India: "Indian business groups are collections of publicly traded firms in a wide variety of industries, with a significant amount of common ownership and control, usually by a family" (p. 867).

opening access as a way to protect those interests. In Britain, the shift occurred after a period of toleration during which many groups that were unable to secure corporate charters formed joint-stock companies that brought them most of the benefits of incorporation. When court decisions threatened these only semi-legal organizations, pressure grew to open the system up. In Massachusetts, change occurred in a context where different groups were competing for political power and, whenever they won, gained the right to charter organizations that benefitted their own members and close down or fetter those their opponents had previously formed. In other words, in both contexts there were prior developments (the advent of laissez-faire attitudes toward business organizations in Britain and the emergence of quasi-democratic political competition in Massachusetts) that changed the interests of the elites in ways that facilitated the achievement of open access without in any way making it an inevitable outcome. These prior developments are associated in British and U.S. history with the transformations of the Age of Revolutions, but the indirect nature of the path from these big events to the achievement of open access raises questions about what other sorts of contexts might alter the interests of elites in similar ways. These questions can be framed in conceptual, empirical, and historical terms.

Finally, what is the nature of the open-access social order? Why does it appear that allowing open access to organizations contributes to economic growth and political democratization in a mutually reinforcing way? Why does open access seem to be self-sustaining? That is, what properties make it an equilibrium outcome?

In limited access societies the benefits that elites derive from continuing to restrict competitors' ability to form organizations are clear. Wallis (2011) explores the more difficult question of what individuals and organizations might have to gain from preserving open access. At the level of the individual, he hypothesizes, people are willing to support the widespread

provision of organizational tools (rules) if the rents they receive from their own organizations are small relative to the value to having many organizations. When individuals find that their talents are valued by other organizations, the rent they obtain from their membership in specific organizations decline, even as the value of their human capital rises. If increasing the number of organizations increases the value of an individual's outside options, in other words, it may also increase the incentive to support rules that create more outside options. To give an example, the basketball player Kobe Bryant makes about $30 million a year playing for the Los Angeles Lakers. The rent he receives from playing *for the Lakers*, however, is much much smaller than the rent he receives from playing professional basketball. There are many teams willing to pay him about $30 million to play for them. Is it in Bryant's interest to break the rules—cheat—in an effort to create an advantage for the Lakers? If he cheats, he puts the entire NBA at risk. Bryant stands to lose much more from damaging the NBA than he gains from promoting the interests of the Lakers. Most of the rents he receives are from being a professional basketball player, rather than from being a member of the Lakers team.

At the level of organizations, the interests that sustain open access must be very different because, as we have just seen, individuals' gain from outside options comes at the expense of their organizations. Wallis argues that what organizations get from the new order is the promise that they can always reconstitute themselves in the future. Under open access, groups cannot be punished with the loss of privileges requisite for their existence. Nor if they lose political power or otherwise fail do they necessarily disappear. Wallis posits that, when the continuation value of being able to fail and reform in the future increases relative to the value of maintaining a specific privilege (and foregoing the possibility of reforming in the future), organizations will find it in their interest to support impersonal rules, even at the cost of lower rents in the present.

Of course, it is difficult to make the case that elites will understand their long-run interests, let alone the case that at reasonable discount rates the gains from the being able to reform in the future outweigh the benefits of restricting access in the present. Other scholars, therefore, take another tack and explore the ways in which the spread of organizations itself creates the conditions that sustain open access. Robert Post and Nancy Rosenblum (2002) summarize the three main mechanisms by which, according to this literature, civil society perpetuates itself. Organizations, first of all, discipline the government by providing focal points for resistance against the abuse of power. Second, they serve as vehicles for coordinating participation in the democratic process. Third, they are acculturation devices that help imbue the population with democratic values. To this list, Jacob Levy (forthcoming) adds a fourth mechanism—that the spread of organizations facilitates an ideological tipping toward a pluralist version of liberalism that takes the first three points as axiomatic.

Levy's work suggests that the pluralist variant of liberalism (the belief that allowing many diverse organizations is good for society) never entirely vanquishes the alternative view that organizations are dangerous vehicles of particularism that should be eradicated or contained. From this perspective, it would be particularly useful to explore the histories of societies (for example, those on the European continent) where the late nineteenth century brought open access to business forms like the corporation but not to the other organizations of civil society (Guinnane 2010). It would also be useful to look more closely at the canonical cases of Britain and the U.S. to get a better sense of the movement toward open access—the extent to which it was uneven and/or contested and how these patterns may have mattered. As Bloch and Lamoreaux (forthcoming) have found for the U.S., the spread of corporate form from beyond businesses to other kinds of organizations was neither automatic nor an alloyed benefit. George

Washington's farewell address in 1796 identified two major threats to American democracy: political parties and organized economic interests (Wallis 2008). Fifty years later, the United States had the first organized mass political parties in human history, as well as the first widespread general incorporation laws giving economic interests the ability to organize at will with tools that the government supplied. Perhaps it is not surprising that Americans in the 1840s were deeply schizophrenic about whether organizations were good or bad for the republic. But they are still schizophrenic today. The recent furor over the Supreme Court's ruling in the *Citizens United* case is only the most recent manifestation of the conflict between our fear of organized interests and the idea that allowing organizations free range for their activities promotes the social good.

The Conference

Our plan for the September 14 planning meeting is to start the day with a discussion of the ideas in this paper and then take up each of the three sets of questions in turn. In preparation, we would ask each of you to read and think about our paper and prepare some brief remarks (five to ten minutes, at most) about what kind of paper you might contribute to a subsequent meeting (or, if you will not be able to contribute, what kind of paper you think needs to be done). The schedule for the day will be:

8:30-9:00	Continental Breakfast
9:00-9:15	Opening remarks by Lamoreaux and Wallis
9:15-10:15	Discussion of Lamoreaux and Wallis paper
10:15-10:30	Coffee Break
10:30-11:30	Discussion of Organizations (Dixit, Greif, Gibbons)
11:30-12:30	Discussion of Transition (Hilt, Bloch, Brooks and Guinnane, Levy)
12:30-1:30	Lunch
1:30-2:30	Discussion of Open-Access Society (Weingast, Zuckerman Silvan, Levi)

2:30-3:00 Wrap up

We look forward to hearing your ideas!

The following list includes all of the people we have invited who have indicated they want to be involved in the project. Some people will not be able to make the planning meeting on September 14, but will be at the following meeting in the spring.

Organizers

Naomi Lamoreaux, Economics and History, Yale University & NBER
naomi.lamoreaux@yale.edu

John Wallis, Economics, University of Maryland & NBER
Wallis@econ.umd.edu

Participants:

Ruth Bloch, History, UCLA
bloch@history.ucla.edu

Nicholas Bloom, Economics, Stanford University & NBER (won't attend the planning meeting)
nbloom@stanford.edu

Richard Brooks, Yale Law School
richard.brooks@yale.edu

Avinash Dixit, Economics, Princeton University
dixitak@Princeton.EDU

Robert Gibbons, Economics, MIT & NBER
Rgibbons@mit.edu

Claudia Goldin, Economics, Harvard University & NBER
cgoldin@harvard.edu

Avner Greif, Economics, Stanford University
avner@stanford.edu

Tim Guinnane, Economics, Yale University
timothy.guinnane@yale.edu

John Haltiwanger, Economics, University of Maryland & NBER (won't attend the planning meeting)

Haltiwan@econ.umd.edu

Eric Hilt, Economics, Wellesley College & NBER
ehilt@wellesley.edu

Margaret Levi, Political Science, University of Washington
margaret.levi@gmail.com

Jacob T. Levy, Political Science, McGill University
jtlevy@gmail.com

Walter W. Powell, Education, Stanford University (won't attend the planning meeting)
woodyp@stanford.edu

Nancy Rosenblum, Political Science, Harvard Univeristy (won't attend the planning meeting, may attend in the spring)
nrosenblum@gov.harvard.edu

Raffaella Sadun, Harvard Business School & NBER (won't attend the planning meeting)
rsadun@hbs.edu

Ezra W. Zuckerman Silvan, Sloan School of Management, MIT
ewzucker@MIT.EDU

Barry Weingast, Political Science, Stanford University and Hoover Institution
weingast@stanford.edu