
Development Traps II: Frozen Accidents

“Jorge Luis Borges wrote a story about somebody who made a model of alternative possible histories of the universe, in the form of a garden of forking paths... [T]he history that is actually seen is co-determined by the simple fundamental laws, and by an inconceivably long sequence of accidents or chance events, which can come out in various ways and in advance one can predict only the probabilities of the different outcomes... some of these accidents produce a great deal of regularity in the future. And those we can call frozen accidents.” Murray Gell-Mann¹⁰³

8.1. Introduction

In the last chapter, we studied complementarities. We’ve seen how such externalities may precipitate a coordination failure. The main idea is simple but powerful: there are many situations in which the taking of an action becomes more profitable when others take similar actions. That could generate multiple equilibria: individuals (or different sectors of an economy) could get locked into suboptimal choices, when there is another configuration that makes agents better off and is *also* self-reinforcing.

This chapter retains some of those notions, but switches the focus. In Chapter 9, the multiplicity of outcomes stems from *beliefs* about what others will do. But beliefs are only one of many things that could be influenced by historical experience. Among other things, think of savings or fertility behavior, institutions such as colonization, cultures such as gender attitudes or traditional norms, past methods of production, political arrangements, and yes, beliefs too — the pre-existing and perhaps pervasive presence of a particular self-reinforcing equilibrium.

The footprint of History may be light, or it may be heavy. An example of the former is fashion, or currency speculation. A fashion currently in vogue, or the holding of

¹⁰³Interview, December 2001, <https://www.nobelprize.org/prizes/physics/1969/gell-mann/interview/>

a currency in an emerging market, can change overnight: if people expect a change, that very expectation can be self-fulfilling. While no extreme view is entirely realistic, it is almost as if History has no role to play at all in such matters. The light footprint manifests itself as equilibrium *multiplicity*: several outcomes are possible at any date. Some of that is true of dating platforms, revolutions, or economic crises, but here the weight of History is somewhat heavier, as described in the previous chapter. The harder it is to change some relevant "state variable" — the capital stock, technological platforms, social norms, political systems, institutions — the less likely is it that society today can nimbly switch from one equilibrium to another. Such situations are therefore less likely to exhibit multiple equilibria. But the multiplicity hasn't gone away by any means. It is reincarnated in "history-dependence," in the fact that historical experience — and even historical accident — can have persistent effects as we move into the future. Multiple equilibria are replaced by multiple *steady states*.

In this chapter, we explore some of the ways in which the Past might cast its shadow on the Future. But before we begin, we remind ourselves again of the limits to the explanatory power of this theory. The theory predicts certain *subsets* of outcomes: the ones that lock together into self-fulfilling equilibria or steady states. But it does not say much about the *particular* combination that one might expect to see in certain settings. A search for ultimate explanations as to why, say, one part of the world was doomed to underdevelopment and another part was not, is typically dominated by all sorts of hindsight-based arguments. We should regard those with a (large) pinch of salt.

Here's a little parable. Imagine there are two types of occupations: skilled and unskilled. Both types enter as essential inputs into production. A large number of identical, altruistic parents must make the decision to "skill" their children at a given cost. Is it possible for all of them to make the same *choices*? The answer is no. If all of them leave their descendants unskilled, then the return to skilled labor will become enormously high, encouraging the acquisition of skill. Yet, if all parents were to invest in their children: the skill premium would vanish or even turn negative. Therefore *identical parents must make different decisions*.

The skill premium adjust to make this possible. Being identical, these parents must be indifferent in their choices. Parents with skilled children enjoy the pleasure of having skilled offspring, while the parents of unskilled children enjoy higher consumption. But of course, the children — who had no say in those choices — do have a definite preference: they would strictly prefer to have been educated.

Now suppose we, as economic detectives, set out to find just why some particular children were skilled and others were not. In each case, we could use *ex post* reasoning — oh, their parents were "more" altruistic, or a teacher encouraged the child, or the child goofed off at school — maybe we could parse many life histories in this way with the benefit of hindsight. But in the abstract world of the parable itself, all the children and their parents were identical to start with, and *yet* their lives diverged. There may be an answer to the question: must we observe certain *aggregate* combinations of historical fortune in society? But there may be no predictable answer to the question of just who will occupy which space in those combinations. These are accidents, and — in the words of the great physicist Murray Gell-Mann that begin this chapter — they are *frozen accidents* that cast long shadows into the future. For the *children* of the

educated children will surely benefit from the accidental event of their parents having been educated, and the children of the uneducated will suffer.

8.2. History and Economic Growth

Recall that the very idea of convergence negates the past. Convergence asserts that if there are historical shackles, they will be thrown off as different countries come together, barring so-called “exogenous parameters” such as savings rates, reproductive proclivities, or culture. Convergence is based on the idea of diminishing returns to inputs: as capital per-capita rises, its marginal product falls, so that poorer countries grow faster than rich countries. But diminishing returns is not gospel, and so-called exogenous parameters may not be exogenous at all. Reversing these notions opens the door to history-dependence, or the persistence of long-ago, possibly accidental events.

Figure 8.1 depicts a particularly stark version of increasing returns, generated by the fixed cost of technology adoption. The left panel shows us two production functions, each using per-capita capital k as input, just as in Chapter 3. The first production function has the usual concave shape with diminishing returns to k . The second production function comes at a price. It requires a certain outlay of capital or a “sunk cost” shown by s — perhaps an investment in entirely new machines, or fresh land acquisition, or the training of personnel — before *any* output is produced at all. But then output rapidly rises until it outstrips the old production function, overtaking it at the per-capita capital level x . If you were an entrepreneur with access to both these functions, your technology would be the outer envelope of these two functions, as shown by the thick outer shell. In effect, *this* is your production function f . Notice that it does not exhibit diminishing returns to capital, or at least not everywhere. The output-capital ratio falls and then rises (and later falls again), as shown by the curved arrows in Panel A.

Panel B plots exactly the same graphical device that we used to solve the Solow model; see Figure 3.4. The difference is that the line $(1 - \delta) + s\Gamma(k)$ inherits the movement of the output-capital ratio Γ . So this line could well cut $(1 + n)$ *more than once*, as the diagram emphasizes.¹⁰⁴ But that gives us three candidates for steady states in the Solow model — the points k_1^* , k_2^* and k_3^* at which the two lines intersect — not one! Now initial conditions matter a great deal. Starting from any initial capital stock larger than k_2^* , the system will converge to the stable steady state k_3^* , using arguments that we’ve seen more than once. But if the

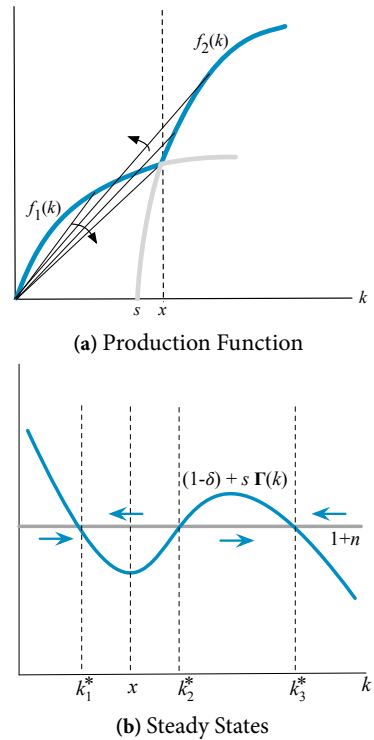


Figure 8.1. The Growth Model With Increasing Returns

¹⁰⁴Under pure diminishing returns, this possibility of multiplicity would be categorically ruled out.

initial capital stock is located anywhere below k_2^* , the system will converge to k_1^* .¹⁰⁵ Thus two societies with the same “economic DNA,” in the sense of having all their parameters and technologies identical, could converge to very different steady states, by virtue of the fact that they start from different initial capital stocks.

There isn’t an easy bridge from k_1^* to k_3^* ; the unstable steady state k_2^* stands in the way. With capital just above k_2^* , the economy is propelled to the higher steady state at k_3^* . But to the left of k_2^* , just the opposite is true: the system retreats to the low-level trap at k_1^* . A small “accident” around k_2^* can freeze into a persistent long-term effect.

8.3. Historical Barriers to Entry

Of course, it isn’t just in the growth model that fixed costs (or increasing returns more generally) opens the door to frozen accidents. There are many other instances.¹⁰⁶

Historical barriers to entry are one of them. Figure 8.2, Panel A, illustrates the predicament of a local car manufacturer of cars. There are two average cost curves for car production depicted in this diagram. The higher curve pertains to an incumbent producer, perhaps a multinational operating in a local market. The multinational is producing quantity $\bar{Q}$, with average cost c . It’s selling the cars at price p and is making a profit.

The lower average cost curve belongs to an ingenious local producer, Zayira, who has designed a new model particularly suitable to local conditions. This is fuel-efficient hybrid that looks very good, because Zayira has an eye for both technology and design. To show both on the same diagram, I’ve translated Zayira’s better design into a lower cost curve: each “effective unit of car” can be produced more cheaply. Notice that both average cost curves are downward-sloping in the figure. The idea is that there is increasing returns to scale in car production — the larger the market, the lower the average cost.

Notice that Zayira could easily enter the market with her new car if consumers were to *instantly* switch loyalties. She could sell quantity $\bar{Q}$ from day one and undercut her higher-cost rival with ease. But Rome wasn’t built in a day and neither are new products. More likely the new car will have to filter its way into the market, with more and more people switching over time. But this gradual transition leads to the following dilemma: *during the transitional switching period, Zayira must function at a loss*. This is easy enough to see from Figure 8.2. If the going multinational price is p , Zayira must undercut this

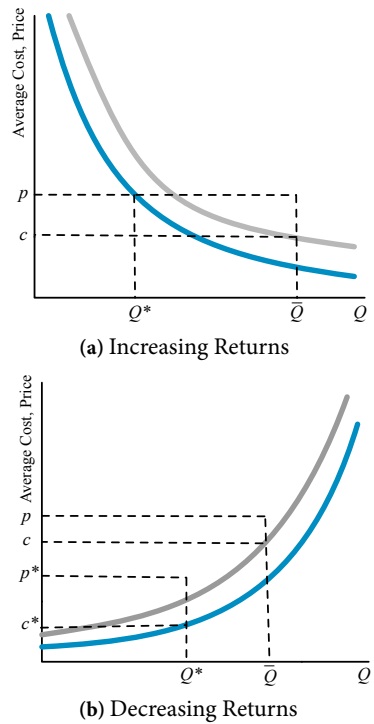


Figure 8.2. Returns to Scale and Barriers to Entry

¹⁰⁵Note that $k_2^* > x$, the stock at which the crossover to the new technology occurs. So for certain initial capital stocks, the economy could wind down on the new technology, even if it were using it to start with.

¹⁰⁶See the classic paper by Allyn Young (1928) for a discussion of increasing returns and market size.

price from day one if the transition to her product is ever to get under way. At least until production levels reach the point Q^* , she will have to contend with losses. Faced with that prospect, it is entirely possible that Zayira may be dissuaded from entering the market in the first place.

Three things have caused this paradoxical problem (paradoxical, because the better design may lose). First, unit costs are high when production levels are small. This is at the heart of those nasty transitional losses. Second, credit markets are missing or incomplete. If capital markets were perfect, Zayira could get a loan to cover her interim losses. Otherwise, and in the absence of deep pockets to self-finance the losses, the whole (socially worthwhile) project may be scuttled.¹⁰⁷ Finally, customers switch slowly, so that it takes time to build a new market, even if that market is finally profitable.

Note that in the absence of increasing returns to scale, the other features pose no problem at all. To make sure you understand this, consider Panel B of Figure 8.2, where just one thing is different: I've replaced increasing returns by decreasing returns, so that average costs are now increasing in scale (or perhaps constant). It is still true that Rome can't be built in a day, but now it doesn't matter: Zayira can invade the market at a small scale Q^* (see Panel B), charge p^* , and *make a profit from day one*. These profits will increase steadily as more and more customers adopt the new product.

We can take this argument a step further. If it is perceived that new technologies cannot invade existing markets for one or more of the above reasons, the incentives to develop those technologies can be crippled. This has far-reaching implications, for it suggests that the technology frontier will be highly clustered across countries. Newer, more appropriate technologies that take advantage of local conditions in developing countries may be extremely difficult to engineer and bring to market.

8.4. Market Size

The argument just made shows entry barriers might arise, even in existing markets dominated by inefficient producers. Worse still, if *many* activities are restricted in this fashion, market size itself will be affected, an argument that dates back at least to Young (1928), though we rely on the formulation in Ciccone and Matsuyama (1996).¹⁰⁸

One feature of economic development is the creation and use of increasingly sophisticated methods of production, often characterized by their "roundaboutness." Almost any productive activity can serve as an example. For instance, in developing countries, construction can be a pretty labor-intensive activity. The area is cleared by hand, rubble is removed in small baskets carried by hand, cement is often mixed at the site, and walls are put up brick by brick. In industrialized economies, many of these tasks have been automated: cranes are used for clearing and prefabricated walls are erected at the site. Each such input, in turn, is produced through a complicated activity: think about how cranes and prefabricated walls are themselves produced. Thus the final production of a house is reduced to a large series of steps, each of a high degree of sophistication and requiring the provision of many intermediate inputs.

¹⁰⁷ Just *why* capital markets might be missing is a subject of discussion elsewhere (see, for instance, Chapters 11 and 19).

¹⁰⁸ For other studies that use similar formulations, see Dixit and Stiglitz (1977), Grossman and Helpman (1991), or Rodríguez-Clare (1996).

These sophisticated inputs can be extremely costly to produce if they going to be sold in tiny markets. The manufacture and sale of cranes requires that there be a fairly large demand for cranes in construction, and so it is with prefabricated walls. Otherwise it is simply not worth setting up separate plants to manufacture these items. In other words, intermediate inputs are often produced under conditions of increasing returns to scale, and if market size is small, those activities may not be profitable. The twist in the tale is that the absence of such inputs feeds back on market size, in what can be a vicious cycle.

To see this in concrete terms, suppose that output is produced using a constant returns to scale technology that includes as inputs intermediate goods as well as labor. Specifically, suppose that the production function is given by

$$Y = [x_1^\alpha + x_2^\alpha + \dots + x_n^\alpha]^{1/\alpha},$$

$0 < \alpha < 1$ and $(x_1, x_2, \dots, x_n)$ are quantities of intermediate inputs via n intermediate processes as in the constriction example. Think of n as the number of roundabout intermediate steps in producing the final output. Each of these inputs requires substantial costs to create and implement, but suppose that once that's done, each unit of each intermediate can be produced by one unit of capital. Dividing that capital among the n inputs, we see that the "indirect production function" is given by

$$Y = [n(K/n)^\alpha]^{1/\alpha} = n^{(1-\alpha)/\alpha} K.$$

This equation shows you, clear as day, why roundaboutness can be productive. If each input is subject to diminishing returns, so that $\alpha < 1$, then more input variety can help to combat such diminishing returns, effectively raising the productivity of raw capital in creating final output. Mathematically, the term $n^{(1-\alpha)/\alpha}$ is akin to productivity, and it is increasing in variety n .

But here's the catch. Inputs don't come for free. They need to be set up and each such change is a costly modification to the production process. If market size is small, n will be small as well, because the reward to those modifications will be limited. That lowers productivity because of the argument in the previous paragraph and generates low income in the economy. Low income in turn generates a low demand for the final good, and the vicious cycle is complete. The other side of that coin is a virtuous circle, to be sure, but the vagaries of history can determine and pin down a particular outcome.

In short, input variety and market size are determined jointly, each feeding on the other. This is a complementarity of sorts, on a grand economy-wide scale, and can lead to frozen accidents.

8.5. The Politics of the Status Quo

It is a fair bet that no matter where you live, your neighborhood has been confronted with the prospect of a large infrastructural change — a new supermarket, a highway expansion, the building of a new subway line, and so on. None of those changes come for free. Many years ago, two streets ran through the location of the New York apartment block where I live, with its accompanying residences and shops. Now, as I write, many changes and redevelopments are planned in my neighborhood, with uncertain consequences. These changes came — and will come — with winners and

losers. But many of us can't be quite sure whether we will be winners or losers, and that very uncertainty can create an enormous bias towards the status quo, and a resistance to change. The accompanying box on dam-building tells you that by no means is such change restricted to urban areas or as matters as relatively localized as a grocery store. The repercussions can be enormous.

The bias of History shows up in all sorts of ways. One central form is the aura of legitimacy conferred by the status quo. If something is in place already, the existing beneficiaries acquire a moral right, and sometimes that sense of morality can be upheld by the law. In this sense the losers from change always appear to possess asymmetric power, especially since the winners are yet to be seen — after all, the policy has not been implemented yet. But it is also the case that the potential winners are very well-organized — after all, they are possibly responsible for bringing the project into being to start with. The resulting titanic struggle between vociferous losers and shadowy, often publicly silent winners, can be seen again and again, and this asymmetry is profoundly determined by the historically given status quo. Matters can go either way.

As one core way of settling a policy issue, consider a referendum based on a majority vote. Suppose that under a proposed policy, it is known that a fraction n of the population will be beneficiaries, each standing to gain \$1, while the remainder will lose \$1 each. If $n > 1/2$, then the policy is a “good one” in the aggregate.¹⁰⁹ Imagine that policy is debated until the above statistics are made clear, and then voted upon.

First, if everyone understands perfectly whether they will be winners or losers, then the policy stands or fails on its merits, and the historical status quo plays no role. If $n > 1/2$, the policy will be voted in, and if $n < 1/2$, the policy will not see the light of day. But things are not that simple. People understand the status quo, but the returns from something entirely new are uncertain. It may be that everyone understands the *statistical* impact: that there will be n beneficiaries, but not their identities. In that case the expected payoff from the policy is

$$n * [\text{payoff from a dollar gained}] + (1 - n) * [\text{payoff from a dollar lost}],$$

which is positive if $n > 1/2$, and negative if $n < 1/2$. So once again, the better alternative wins, and History plays no role in determining the outcome. But this argument presumes that no one is risk-averse. For instance, think of a move to unequivocally free international trade. Will engineers or doctors benefit? Will textile workers or agricultural laborers lose out? The answers are difficult and complex: they depend on the sectors that will succeed in the international market, and the sectors that die out. In the presence of such uncertainties, *a reform policy that is commonly known and agreed to be beneficial ex-post may well be vetoed ex-ante*. The payoff to a dollar gained may be lower than the loss from a dollar lost. Because the status quo is given by History and is known, uncertainty generates status quo bias.

Fernández and Rodrik (1991) show that this bias extends in a subtle and interesting way, even when all voters are risk-neutral (so that a dollar gained or lost is valued to the same degree). They retain the idea that some people don't know their winner-loser identity, but presume that some others *do* in fact know. Think of these as the winners who are pushing the policy to start with. This scenario is a combination of the two extremes discussed earlier. The fraction (call it k) of *known* winners will, of course,

¹⁰⁹ For the purpose of this example, we neglect the (very real) concern of distributional impact.

vote in favor. But the remainder do not know whether they will be winners or losers. They only know that their chances of being a winner are given by $(n - k)/(1 - k)$. After all, they know that they are *not* among the k who *do* know, so they are among the uncertain group of size $1 - k$. All they know is that there will be $n - k$ winners among this group. So their expected payoff from the policy is

$$\frac{n - k}{1 - k} * [\text{a dollar gained}] + \frac{1 - n}{1 - k} * [\text{a dollar lost}],$$

but this could well be negative even if the majority are winners! Indeed, that will happen whenever

$$\frac{n - k}{1 - k} < \frac{1}{2}, \text{ or } 2n - 1 < k.$$

But that has the odd implication that whenever $2n - 1 < k < 1/2$, the policy will fail to pass even if it is known among the voting public that a majority will stand to benefit!

This analysis presumes that some of the winners know they are winners in advance, which (conditional on the overall fraction of winners) actually *reduces* the chances that the others — who do not know — will be winners as well. The argument would be flipped if some losers knew their fate to start with. This increases the chances for a currently uncertain person to be a winner, thereby generating an “anti-status-quo” bias. That may well happen in many cases. But remember: *someone*, possibly an organized group, suggested the policy in the first place, and did so because they *knew* they would benefit. That increases the chances that more winners than losers will know their identity in advance, which leads to the status-quo bias described above.

Institutions, Policies and History: *Narmada Bachao*

All new policies have winners and losers. Whether those policies are implemented will depend on whether political power lies in the hands of those who stand to gain from change. Sometimes those changes are, on balance, for the better. Sometimes, with the rise of newly minted economic power, the answer is more ambiguous.

Consider the intense controversy over the construction of the Sardar Sarovar Dam on the Narmada River in India. The foundation stone for that project was laid by Prime Minister Nehru in 1961. But construction did not start until a quarter of a century later, and the protests were sustained, long-lasting and at times explosive. Members of a dedicated movement—the *Narmada Bachao Andolan*^a—have protested the construction by threatening to drown themselves in the waters of the river. The World Bank, which had offered funding for the dam, pulled out in the 1990s. It was widely perceived that building the dam (and connected irrigation canals) would displace hundreds of thousands of people. But of course, there would be beneficiaries too: numerous downstream districts receiving controlled irrigation, as well as recipients of electricity in multiple Indian states.^b

When faced with benefits and costs that have uneven impact, should we just tot them all up and figure out if the *net* benefit is positive? Some of the calculations aren't hard to do. For instance, it should surely be possible to estimate the loss of land and crop output from the dam and the losses to displaced villagers, as well as the potential gains from surrounding areas to be serviced by the dam. But other calculations are far harder. How does one evaluate the loss of an ancestral home to a farmer? A particularly cynical answer might be to use the market price of his land, but a market price for land that threatens to be underwater pretty soon may not be ...er, adequate. Likewise, it may also

be difficult to estimate benefits, and the issue of proper valuation is compounded by strategic considerations. Cost–benefit analysis is a dangerous “science.”

But even if we knew the benefits and the costs, there is still the question of whether the winners will compensate the losers. That promise may not be credible. If that credibility is missing, perhaps because of a missing institution to enforce compensatory payments, the efficient rules will not be written in the first place. That applies to projects such as dam-building: it may be impossible to tax all the potential beneficiaries until *after* the project has been carried out, but once that’s happened, we are politically (although not morally) in a new status quo, where the project is done, completed. At that point, compensation may never be forthcoming unless there are institutions that guarantee it. And if that possibility is (rationally) anticipated, then the question of adding up all gains and losses becomes irrelevant. It will all depend on whether and how the historical status quo is respected. It will depend on how hard the losers and gainers lobby on either side. It becomes a question of power rather than the computation of aggregate benefits.

In a ceremony on September 17, 2017, Indian Prime Minister Narendra Modi declared the construction of the Sardar Sarovar Dam complete. In a speech for the occasion, he described “a massive misinformation campaign” by opponents of the dam, blamed the World Bank for its turnaround, and noted that even temples in Gujarat had come forward with funds for the dam. Yet, even as Modi spoke, a symbolic protest was held by activists half-submerged in water, evoking the plight of many households to be displaced by the rising waters of the Narmada.

^a A literal translation is Movement to Save the Narmada.

^b Duflo and Pande (2007) describe this conflictual imbalance of benefits and costs.

8.6. Institutions and History

Perhaps history casts its longest shadow in the shaping of economic, social and political *institutions*, that persist and constrain (or possibly enable) the future development of a country. Remember that so many developing countries today were colonies of the great European powers, and in an attempt to govern those countries — presumably for their own benefit — the colonizers laid down not just infrastructure and capital, but rules and regulations. Sometimes, these regulatory systems were laid over informal norms that had been in place before the colonial era, and sometimes the informal norms were simply eradicated (occasionally with the eradication of the local population as well). It would be extraordinary if these spiderwebs of ambient rules for conducting socio-political-economic life did not continue to matter today. And indeed, they do.

8.6.1. A Word on Institutions. Institutions come in all flavors. There are legal institutions that protect property rights or guarantee that contracts will be enforced, or prevent crime. There are economic institutions that provide insurance against a banking crisis, or enable the public to hold financial stakes in companies. There are political institutions that determine how a government is to be elected or operated. There are socioeconomic institutions that determine pensions in old age, or welfare payments, or unemployment insurance. And last but not least, many institutions only exist in the guise of *norms*. For instance, there are norms that govern good or bad conduct, and codes of reciprocity and sanctions in the event of compliance or deviance. There are traditions that impinge on such diverse matters as crop choice, marriage, contraception, or attitudes to entrepreneurship or moneylending. There are informal networks of relationships that influence the flow of information, or the ability

to monitor the activities of others, and these in turn influence the ease with which certain economic transactions can be implemented.

"Institutions" is a very big word indeed — but often used quite cavalierly in small economic models. And the creation of institutions is also a complex business. Apart from the occasional well-thought-out writing of a Constitution (and perhaps not even that), institutions have always been set up to benefit their creators first and the general population second. Often that's fine: think of institutions that reduce crime, or guarantee the right to vote, or provide infrastructural benefits. Good economic institutions can promote investment and growth: their creators and populaces need not always be at odds. But sometimes the maximization of creator benefits may not coincide with the interests of the general public, and more importantly it may not wash well with large sections of *future* generations. Indeed, bad institutions (such as autocracies) can self-replicate, or morph into worse institutions — even dictatorships — as their original beneficiaries struggle to hang on to their positions of power.

8.6.2. A Few Acres of Snow. In influential work, Kenneth Sokoloff and Stanley Engerman (1997, 2000) have argued that endogenously created institutions can explain the divergent paths taken by North and South America to the present day. In the 18th century, North America did not come close to the lush wealth of the Caribbean, and Voltaire's sneering reference to Canada as "a few acres of snow" wasn't as funny then as it is now. That remark was made in the context of the Seven Years War between the British and the French. "The victorious British were later to engage in a lively public debate over which territory should be taken from the French as reparations — the Caribbean island of Guadeloupe (with a land area of 563 square miles) or Canada" (Sokoloff and Engerman, 2000).

South America had huge mineral riches and lots of native labor. Under their Iberian colonizers, they turned into largely extractive or plantation-based economies. These provided enormous surpluses to a relatively small elite, who owned the bulk of the resources. One would imagine that the high rates of return so ostensibly on display would lead to a high inflow of colonizers from the Old World or elsewhere, driving the super-high profits down to competitive levels. Well, that was not to be the case, and for two reasons. First, the *rights* to those sources of surplus were controlled by legal fiat. For instance, there were strict restrictions on migration to the New World. And apart from these blinkers on commerce, there were constraints placed on political participation. You needed to own substantial amounts of land in order to vote. But the second reason was, in a way, still more fundamental: the nature of economic activity — extraction or plantation production — was essentially large-scale. You had to have substantial wealth to finance your entry on these terms into the New World, just as in the barriers-to-entry example in Section 8.3. And this made it easier to implement the legal and political institutions that I've just mentioned. If ever there was a case of an unequally situated elite, which tried to hold on to power, this was it.

Meanwhile, the US and Canada presented a different aspect. Native labor was not used in production, which was relatively small-scale, with immigrants working the land themselves. In general, the agricultural climate did not lend itself to large-scale plantation production (sugar and later cotton), except in the South, but even here, the size of the sugar plantations was relatively small. Further north, the production of wheat and similar crops meant that productive activities were possible with abundant land

and relatively little capital. This lent itself well to individual, self-cultivating ownership. It was therefore far harder to promote — and obtain support for — institutions that demanded huge wealth requirements, and perpetuated unequal power. It was possible to come in under the radar as it were, and start a life.

In short, the initial mode of production was conducive to accompanying institutions that were more inclusive, democratic and progressive. This isn't to say that it was all a bed of roses in North America,¹¹⁰ but relative to institutional developments in the South, it wasn't bad. This led Sokoloff and Engerman (2000, 228–230) to their provocative and insightful conclusion:

“These early differences in the extent of inequality across New World economies may have been preserved by the types of economic institutions that evolved ... [S]uch biases in the paths of institutional development likely go far in explaining the persistence of inequality over the long run in Latin America and elsewhere in the New World.”

8.6.3. Testing for the Influence of Early Institutions. Do early institutions affect modern institutions, and in so doing, affect economic performance today? As you might imagine, *empirically* answering a question like this is extremely difficult, and any attempt to do so will have its shortcomings. In an important paper, Acemoglu, Johnson and Robinson (2001) [AJR] emphasize the institutional protection of property rights. Specifically, they consider “protection against expropriation,” an index published by *Political Risk Services*.¹¹¹ It is very easy to see that “protection against expropriation” is significantly correlated with per-capita GDP: simply run the regression

$$y_i = C + \beta R_i + X_i \gamma + \epsilon_i, \quad (8.1)$$

where y_i is log per-capita GDP of country i , X_i is a vector of country-specific characteristics, and R_i is the protection measure from *Political Risk Services*. The resulting fit is shown in Figure 8.4. The relationship is strong and positive. But there is good reason to doubt this sort of evidence, or more precisely, to doubt that a convincing *causal* implication can be drawn from it. Like many regressions of this ilk, it suffers from important problems that preclude a satisfactory causal interpretation:

1. *Reverse Causality.* Countries with high per-capita income can afford better institutions, so y_i may be driving R_i and

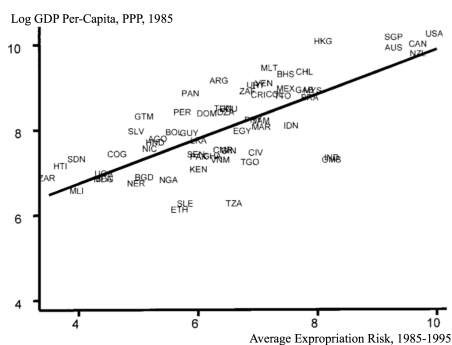


Figure 8.3. Per-Capita Income and Protection Against Expropriation (OLS). Source: Acemoglu, Johnson and Robinson (2001).

¹¹⁰For instance, voting in the United States was largely open to (white) men with substantial property, but the suffrage was rapidly extended in the first half of the nineteenth century. Canada followed in due course, but even at the turn of the twentieth century, the only Latin American country without a wealth or education requirement on voting was Argentina (Engerman, Mariscal and Sokoloff, 1999).

¹¹¹*Political Risk Services* is part of the PRS Group, which makes available risk assessments for businesses seeking to invest in different countries. Without endorsing this company in any way, here is the link for more information: <http://epub.prsgroup.com/the-countrydata-gateway>.

not the other way around.¹¹² In short, the tail could be wagging the dog.

2. *Omitted Variables.* Countries with better bureaucracies or leaders, or a more cohesive “culture,” could provide stronger protection from expropriation *and* do a better job of generating income. Then it’s neither the tail nor the dog that wags each other, but something else (the sight of a chewy bone maybe?) that wags both of them.

There is a third problem that may or may not always be present, but relevant here:

3. *Measurement Error.* The explanatory variable may be measured with error. I’m not talking about deliberate mis-measurement to shift the average observation, but just plain noise that could shift the measured variable equally in one direction or another.¹¹³ Measurement error biases the estimate of β *downwards*. Intuitively, the noisier the measure, the less obvious will it be that it has any effect on any other variable at all.

Reverse causation, or the possibility that some variables may have been omitted, or measurement error in the explanatory variable, all display a common and unifying statistical signature: the risk measure R_i in Equation (8.2) ends up being correlated with the error term ϵ_i . This is the problem of *endogeneity*. Endogeneity biases the estimated coefficient β , and its value is not necessarily indicative of a causal relationship running from R_i to y_i . See the Statistical Appendix for a more detailed discussion.

Of course, AJR are not only aware of these problems but try to do something about it, which is why their paper goes beyond a merely competent compendium of correlations. They propose an *instrument* to deal with these issues. Briefly, an instrument is a third “magic variable” that satisfies two conditions: (i) it is correlated with the explanatory variable that we suspect is prey to endogeneity, R_i in this case, and (ii) it exhibits no further correlation with the dependent variable (y_i in this case) *over and above* the connections via the explanatory variable that is being “instrumented for.”

What constitutes a good instrumental variable? Condition (i) — to be used in the first stage of a two-stage regression exercise — is something you can verify statistically. Just regress the explanatory variable (the one suspected of endogeneity) on the proposed instrument and check out how it works. Condition (ii), known as the *exclusion restriction*, must exclude other pathways connecting the instrument to the dependent variable. You can’t verify this statistically: after all, the instrument *is* correlated with the dependent variable at least via the explanatory variable. You will have to make a *conceptual* case for the exclusion restriction. This is why the hunt of an instrument is something of an art form. It is also — often — an elusive hunt, much like finding a unicorn. Real unicorns might exist, but they are very seldom seen.

AJR use data on the mortality rates of soldiers, bishops, and sailors in different geographical regions in the world (previously colonies of European settlers) to form their instrument. They write: “we hypothesize that settler mortality affected settlements; settlements affected early institutions; and early institutions persisted and formed the basis of current institutions.” Indeed, they are on solid ground here, as it is clear from historical accounts that mortality from diseases such as malaria and yellow fever surely affected the extent of initial settlement, or even the decision to move to a location

¹¹²It may even be that R_i is *perceived* by the analysts to be high where y_i is high.

¹¹³After all, R_i is a simple index that attempts to compress a much broader concept — protection against expropriation, which has many dimensions to it — into a single number.

	Base sample (1)	Base sample (2)	Base sample without Neo-Europes (3)	Base sample without Neo-Europes (4)	Base sample without Africa (5)	Base sample without Africa (6)	Base sample with continent dummies (7)	Base sample with continent dummies (8)
Panel A: Two-Stage Least Squares								
Average protection against expropriation risk 1985–1995	0.94 (0.16)	1.00 (0.22)	1.28 (0.36)	1.21 (0.35)	0.58 (0.10)	0.58 (0.12)	0.98 (0.30)	1.10 (0.46)
Latitude		−0.65 (1.34)		0.94 (1.46)		0.04 (0.84)		−1.20 (1.8)
Asia dummy							−0.92 (0.40)	−1.10 (0.52)
Africa dummy							−0.46 (0.36)	−0.44 (0.42)
“Other” continent dummy							−0.94 (0.85)	−0.99 (1.0)
Panel B: First Stage for Average Protection Against Expropriation Risk in 1985–1995								
Log European settler mortality	−0.61 (0.13)	−0.51 (0.14)	−0.39 (0.13)	−0.39 (0.14)	−1.20 (0.22)	−1.10 (0.24)	−0.43 (0.17)	−0.34 (0.18)
Latitude		2.00 (1.34)		−0.11 (1.50)		0.99 (1.43)		2.00 (1.40)
Asia dummy							0.33 (0.49)	0.47 (0.50)
Africa dummy							−0.27 (0.41)	−0.26 (0.41)
“Other” continent dummy							1.24 (0.84)	1.1 (0.84)
R ²	0.27	0.30	0.13	0.13	0.47	0.47	0.30	0.33
Panel C: Ordinary Least Squares								
Average protection against expropriation risk 1985–1995	0.52 (0.06)	0.47 (0.06)	0.49 (0.08)	0.47 (0.07)	0.48 (0.07)	0.47 (0.07)	0.42 (0.06)	0.40 (0.06)
Number of observations	64	64	60	60	37	37	64	64

Figure 8.4. IV Regression of Per-Capita Income on Protection Against Expropriation, Instrumented by Settler Mortality. Source: Acemoglu, Johnson and Robinson (2001), Table 4. The dependent variable is log GDP per capita PPP in 1995, PPP basis. “Average protection against expropriation risk 1985–1995” is measured on a scale from 0 to 10, and taken from Political Risk Services. Panel A reports 2SLS estimates, instrumenting for “average protection” using log settler mortality; Panel B reports the corresponding first stage. Panel C reports the corresponding OLS regression of GDP per capita on average protection without instrumenting. Standard errors are in parentheses. In regressions with continent dummies, the dummy for America is omitted.

in the first place. (AJR convincingly discuss many examples.) This is their check for condition (i), the “first stage” described above.

The exclusion restriction is less convincing. Recall that this restriction will fail if the instrument is separately correlated with economic outcomes today through another channel. Might the mortality rates of soldiers, bishops, and sailors be somehow correlated with economic outcomes today via a path entirely separate from institutions? One obvious culprit is the disease environment. Malaria comes particularly to mind. The prevalence of malaria would surely have influenced the death rates of these venerable gentlemen (alas, men they almost surely all were in those days). But malaria will also influence current GDP per-capita via its deleterious effects on labor productivity. For lack of space, I refer you to the original paper. The short answer is that AJR attempt to deal with some of these alternative pathways — climate variables, malaria, yellow fever, geography — by simply controlling for them in addition to running the instrumental variables regression. In Figure 8.4, I restrict myself to showing you the basic IV regressions. There are eight of these basic regressions, and they vary according to the exact set of control variables that are used in each case.

In each specification, Panel B tells us the outcome of running the “first stage,” which is a regression of the “average protection” variable today on settler mortality. You can see the important coefficient of this exercise in the first row of this panel. That coefficient is negative and significant, verifying that settler mortality in the distant past is indeed correlated with “good institutions” today — negatively so. The basic idea of an IV regression is then to use the *predicted* values of this first stage exercise as a replacement for the independent variable in the “second stage,” which is in Panel A. The idea is to *only* employ that part of the independent variable (“average protection” in the case at hand) that’s explained by movements in the instrument, excluding other residual movements that could be confounded by concerns of endogeneity. Moving up to Panel A, then, and running this second-stage regression, our coefficients of interest lie along the first row of that panel. In each case, the results are positive and strongly significant. In fact, the sizes of the coefficients are larger than those obtained via ordinary least squares (without instruments), suggesting that if anything, measurement error was the main confound in that case and not reverse causality or an omitted variables problem.

The coefficients are also sizable (apart from being significant). To illustrate, they compare two countries with high and low “average protection,” Chile and Nigeria. The IV estimate in the very first column, which is 0.94, means that a difference of 2.24 in the protection variable is translated into 206 log points in GDP difference, a factor of seven. This is a large difference, but given the two countries, not absurdly so.

I like this exercise because it is far-reaching, ambitious and dares to go out on a limb, linking the distant past to current outcomes and indeed finding a connection. There is much here to criticize — the validity of the instrument among them — but there is a lot to admire as well. That said, one must note that IV regressions of this kind rarely do a good job in identifying the underlying pathway of influence. Yes, settler mortality appears to affect the institutional protection environment today, but *just how* does it do so? The regressions don’t tell us. For that, we could rely on — or at least ponder — the pioneering argument advanced by Sokoloff and Engerman. That would run as follows: high settler mortality led to low settlement densities. These made for monopolies or near-monopolies on initial extraction or production in the colonies. In turn, such monopolies lobbied for prompting restrictive institutions that were conducive more to extraction and expropriation than to socially productive activity.

AJR echo these observations in part as they write:

“Soon after the conquest, the Spanish crown granted rights to land and labor (the *encomienda*) and set up a complex mercantilist system of monopolies and trade regulations to extract resources from the colonies ... There is a variety of historical evidence, as well as our regressions ... suggesting that the control structures set up in the non-settler colonies during the colonial era persisted, while there is little doubt that the institutions of law and order and private property established during the early phases of colonialism in Australia, Canada, New Zealand, the United States, Hong Kong, and Singapore have formed the basis of the current-day institutions of these countries.”

In their pathbreaking work, Sokoloff and Engerman emphasized the chain running from early modes of production to the development of a host of political, legal and social institutions, the repercussions of which (they asserted) are present in the world today. A number of studies follow this route. Dell (2010) tells the story of labor conscription by the Spanish *conquistadores* for silver mining under the *mita* system in Southern Peru and Bolivia. As the silver mines were progressively depleted, *Mita* came to an end in the early part of the 19th century. But the effects of that system are felt to the present day.

As an institution of compulsory labor acquisition, *mita* predates the Spanish arrival. It dates back to the Inca Empire, and roughly speaking it stood for public service from citizens: the mandatory supply of labor for community projects ranging from road-building to irrigation to warfare. The *conquistadores* adopted this system for the Potosí silver mines in 1573 and took it in another direction: to guarantee labor for mining, and thereby assure a steady and growing supply of silver to the Spanish Crown. Gone were the Inca norms, which emphasized reciprocity, community service and a certain degree of social insurance. Under Viceroy Francisco de Toledo, *mita* became a brutal system of coercive labor that demanded one-seventh of the labor force (in certain geographical areas) for work in the silver and mercury mines, and in several other capacities, including but not limited to agricultural labor.

As I've said, *mita* is long gone. But the *mita* districts from which labor was forcibly conscripted are badly off today on a number of counts. Households from these districts are more likely to be engaged in subsistence farming than their counterparts from non-*mita* districts. They are less integrated into road networks. Household consumption is lower on average. The incidence of undernutrition is higher. *Mita* — a system that ended 200 years ago — is correlated with various indicators of poverty *today*. But how would one go about demonstrating such a claim? One natural approach is to run the regression:

$$y_i = C + \beta m_i + X_i \gamma + \epsilon_i, \quad (8.2)$$

where y_i is some outcome of interest for household i (say their consumption), m_i is a dummy variable that takes the value 1 if the household comes from a *mita* district, and 0 otherwise, and X_i is some list of household-specific controls. The idea: if y_i is some index of modern achievement (say, household consumption), then the coefficient β should be negative.

And doubtless, such a finding would be of interest: it would show that the very presence *mita* is correlated with bad outcomes today. But one might want to go further: might *mita* have actually *caused* these present-day outcomes? That more ambitious question must contend with the endogeneity problem. Both reverse causation and the possibility of omitted variable bias create potential pitfalls. For instance, districts with high poverty may have been easier to subjugate, so that “poverty caused *mita*” and not the other way around. Or there could be other unobserved factors correlated with the imposition of *mita*, and *also* with economic circumstances today. For instance, a sudden change of altitude did determine the boundary of a *mita* “catchment area.” But altitude could also independently drive subsequent differences in economic development.^a

The strategy that Dell follows is the exploitation of a *regression discontinuity*. This is a fortuitous situation — for econometricians, that is, not always for the people affected by it — in which there is a sudden change in a policy over a domain with otherwise smoothly varying characteristics. The smoothness allows us to assert that omitted variables, or any factors such as poverty that we would use to build a reverse causality argument, are changing very “slowly” relative to the trauma generated by the boundary break. Any large differences in outcomes must be attributed therefore to that break.^b

In our case the discontinuity comes from the geographical *mita* boundary. There was a limited demand for *mita* labor — after all, there could only be so much labor used in the mines. So *mita* conscription had a boundary, beyond which the population was not subjected to compulsory labor demands. Most of this boundary was demarcated by natural limits — the steep Andean precipice to east and west, as well an uninhabitable salt marsh to the south. The land on either side of such natural boundaries were indeed quite dramatically different. So there would be other “omitted variables” changing along with the transition from *mita* to non-*mita*. But other sections of the boundary were not demarcated by natural transitions. The situation on either side of these zones was reasonably similar — in elevation, terrain, soil quality and rainfall. Social, economic and demographic characteristics were similar as well in the 16th century as *mita* took hold: ethnicity, patterns of settlement, and local tax rates. And so, one might reasonably hypothesize that significant differences we see today on either side of such boundaries could be traced back to a *causal* influence of the *mita*.

With this technique in hand, Dell finds a strong, long-run *mita* effect: *mita* districts close to the boundary have lower household consumption than their counterparts on the other side — by as much as 25% in 2001 by around 25%. The same is true of stunting in children, which is another indicator of living standards. Districts just within the *mita* boundary have anywhere between 10% to 25% higher rates of stunting.

These effects are remarkable: a signature left on present times by a long-dead institution. What could be the channels for such persistence? Dell ascribes a principal channel to the fact that large-scale land holdings — or *haciendas* — could form outside the *mita* boundaries, with the principal landowners providing long-term insurance and stability to the tenants and smallholders who lived: “I hypothesize that the long-term presence of large landowners in non-*mita* districts provided a stable land tenure system that encouraged public goods provision.” For the communally held *mita* districts, there was no effective system of land titling so that all hell broke loose (land confiscation, rebellion, rustling ...).

Hey, but haven’t we heard that large landowners can be nasty and exploitative? The answer is, as many answers often are, contextual: *yes*, in contrast to secure small holdings, *perhaps not* in contrast to anarchy and no holdings at all. “The evidence in this study indicates that large landowners— while they did not aim to promote economic prosperity for the masses — did shield individuals from exploitation by a highly extractive state and ensure public goods.”

^aEither of these situations would cause m_i in Equation (8.2) to be correlated with the error term ϵ_i , so that the estimated β is not necessarily indicative of a causal relationship running from m_i to y_i . See the Statistical Appendix for a more detailed discussion.

^bA regression discontinuity is therefore a special kind of instrument, in which the first stage correlation is sharp and unambiguous (for an instrument in general, this may or may not be the case). The case for the exclusion restriction is typically made by studying the smoothness of other observable factors across the boundary of the discontinuity — see the Statistical Appendix for more details.

8.7. Norms and History

Institutions are but one way in which History can constrain — or enable — future development in fundamentally varying ways. Social norms and attitudes are another. We illustrate this with a story, that begins with the work of the Danish economist Ester Boserup. Boserup made two major contributions to economics. She advanced a theory of population and economic development markedly different from the pessimism

of Malthus; we will have occasion to encounter her work again in Chapter 14. But Boserup was also deeply interested in the interaction between gender and economic development, and sought the roots of gender discrimination in traditional modes of agriculture. She was specifically interested in the idea that the need for physical strength in many agricultural tasks may have driven women out of the agricultural workplace into the home. This would seem to many as an innocuous and perhaps even efficient division of labor. And perhaps it was so, but norms survive longer than modes of production. Or at least this is what Ester Boserup so wisely foresaw.

In her 1970 book, Boserup picked away at the distinction between shifting cultivation and plough-based cultivation. Unlike the former, in which light tools such as the hoe are used and women participate actively, the plough requires strength to pull, or strength to control the animal that might pull the plough. The plough also reduces the need for weeding, which again is a task that is “female-intensive.” Summarizing her work, Alberto Alesina, Paula Giuliano and Nathan Nunn (2013) [AGN] write:

“Societies characterized by plough agriculture, and the resulting gender-based division of labor, developed the belief that the natural place for women is within the home. *These cultural beliefs tend to persist even if the economy moves out of agriculture*, affecting the participation of women in activities performed outside the home, such as market employment, entrepreneurship, or participation in politics.”

(I added the emphasis to underline the argument played by the persistence of beliefs that might transcend the initial mode of production that generated those beliefs.)

AGN set out to test this hypothesis. There is data available on plough use by ethnic group from George Peter Murdock’s *Ethnographic Atlas*, a fascinating compendium of pre-colonial data for over ethnic groups all around the world. We know about plough use for over 1000 such groups, and with a few exceptions, the data is about aboriginal use, before European contact. We also know from this data that societies that used the plough in cultivation *did* employ fewer women in agriculture, in land clearance, soil preparation, planting or harvesting. But of course, all of this data comes from the same traditional period, and says nothing about the effect (if any) of traditional agrarian practices on gender roles today. AGN look at *modern* societies. By studying modern data on land regions (countries or smaller units) and the composition of languages spoken on these units, and by connecting those languages to the ethnicities in Murdock’s *Atlas*, AGN construct an index of “ancestral plough use” for each geographical unit today. It is important for the appreciation of their result that that index may have little or nothing to do with plough use *today* in that unit. For instance, we could construct such an index for New York City by pooling together ancestral plough use for all the ethnicities who live here. (And the last time I checked, there wasn’t much by way of agriculture in New York City.)

Now we can regress *modern* indices of female empowerment on *ancestral* plough use. AGN do so at the country level by using the equation

$$y_c = \alpha + \beta \text{Plough}_c + \mathbf{X}_c^H \Theta + \mathbf{X}_c^C \Pi + \varepsilon_c, \quad (8.3)$$

where y_c is the outcome of interest in country c , and $\mathbf{X}_c^H$ and $\mathbf{X}_c^C$ are historical and contemporary country-level controls. I reproduce a subset of the results in Table 8.1. There are three outcomes that AGN work with: female labor force participation in

	FLP		Female Firms		Female Politics	
Plough Use	-14.90*** (3.32)	-12.40*** (2.96)	-16.24*** (3.85)	-15.24*** (4.06)	-2.52 (1.97)	-4.82*** (1.78)
Hist Controls	y	y	y	y	y	y
Mod Controls	n	y	n	y	n	y
Observations	177	165	128	123	153	144
Adjusted R-sq	0.20	0.37	0.14	0.11	0.14	0.27

Table 8.1. Ancestral Plough Use and Modern Gender Outcomes. Source: Alesina, Giuliano and Nunn (2013, Tables II and III). FLP = female labor force participation, 2000; female firms = share of firms with female ownership, 2003-10; female politics = share of political positions held by women, 2000. Robust standard errors are reported in parentheses. ***= significance at 1%. Historical controls include tropical climate, presence of domesticated animals, and political hierarchies. Modern controls are per-capita income in 2000 entered both linearly and with a quadratic term. The table here reports estimates without continent fixed effects; see original table for a version with such fixed effects.

2000,¹¹⁴ the share of firms owned by women over 2003–2010, and the share of political positions occupied by women in 2000.¹¹⁵

The regression coefficients are persistently and significantly negative, with the exception of female participation in politics when there are no contemporary controls for per-capita income. The variable turns significant when income controls are introduced; we return to this point below. With all controls in place (columns 2, 4 and 6), a one standard-deviation increase in *ancestral plough use* appears to reduce female labor force participation *today* by 11.4%, female ownership *today* by 20% of its sample mean, and female participation in politics *today* by 19% of its sample mean. That is, in itself, quite remarkable.

As in the Acemoglu-Johnson-Robinson paper, one can question these findings on the grounds of endogeneity. There are problems of reverse causation. Societies with bad gender norms could have adopted the plough — and gender norms today could be correlated with gender norms in the distant past, thus creating a correlation without any of the Boserupian causality that we seek. Or there may be omitted variables at work: historically richer countries could have adopted the plough — and historically richer countries are richer today and so have better gender attitudes today. You can see some hint of this in the reduced-form regression of Table 8.1, in the differences in significance between columns 5 and 6. Notice, by the way, that this confounding term goes the other way: it would work to actually *reduce* the negative correlation that we see in the reduced-form regression. But a cofound is a confound.

To get around these issues, AGN search for an instrument. The idea is to look for something else that might affect plough use but at the same time has no separate correlation with gender norms today. One approach is to look at the prevalence of crops that are particularly amenable to the use of the plough. We know that crops such as wheat, barley, rye, wet rice, and teff¹¹⁶ are “plough-positive”: they take well to be

¹¹⁴This is the percentage of women in the labor force expressed as a share of all women of working age.
¹¹⁵This variable is measured by the proportion of seats occupied by women in the national parliament.
¹¹⁶Teff is the main ingredient in the delicious Ethiopian *injera*. If you haven’t eaten *injera*, stop reading now and find an Ethiopian restaurant!

Panel A: Second-Stage 2SLS Estimates			
	FLP	Female Firms	Female Politics
Plough Use	*** -21.63 (5.25)	*** -17.49 (5.53)	*** -6.46 (2.33)
All Controls	y	y	y
Observations	160	122	140

Panel B: First-Stage 2SLS Estimates			
	Plough Use		
Plough-pos	*** 0.74 (0.08)	*** 0.86 (0.08)	*** 0.08 (2.33)
Plough-neg	0.12 (0.12)	0.10 (0.17)	0.13 (0.13)

Table 8.2. Plough Use and Gender Outcomes, 2SLS. Source: Alesina, Giuliano and Nunn (2013, Table VIII). Robust standard errors are reported in parentheses. *** = significance at 1%. All controls = historical + modern as in Table 8.1. **What about the F-stats on the first stage?**

cultivated using the plough. Other crops, such as maize, sorghum, millet, and tubers are “plough-negative,” and are more suited to shifting cultivation.

How about using these crops as an instrument, might that work? The answer is no. They would certainly be correlated with plough use, so the first condition for an acceptable instrument is indeed met. But they could easily violate the exclusion restriction: for instance, countries with bad gender norms might have *chosen* to grow plough-positive crops, thereby placing women in the homes where they were thought to belong. Plough-positive crops are therefore not that much better a variable than plough use itself. The strategy that AGN use instead is geo-climatic suitability of the land inhabited by each ethnic group for plough-positive and plough-negative crops. Specifically, they look at soil, climate and geographical characteristics that are suitable for the plough-positive crops wheat, barley, and rye, and likewise for the plough-negative cereals foxtail millet, pearl millet and sorghum. Once again, we can hope for a decent correlation with plough use.

As for the exclusion restriction: might there be any reason to believe that geo-climatic *suitability* for crops — not the crops themselves — might correlate with gender norms, apart from the specific channel of interest, which is the adoption of the plough? There might be. Geographical or climatic features — such as terrain, temperature, or rainfall — that correlate with crop suitability might also influence gender attitudes more generally. These can be controlled for separately both stages of the IV regressions, taking care of the problem to some degree. It is also possible that ethnic groups with certain gender norms may have traditionally migrated to or chosen to remain in locations that were suitable for crops that used the plough. This is a concern. It is one that we can probably do little about. As I’ve already remarked, the search for the perfect instrument is much like the hunt for a unicorn.

Table 8.2 presents some of the IV estimates from AGN. Not much changes relative to the OLS in terms of sign or significance. As for size, the IV coefficients are larger. This is not surprising. As already discussed, historically richer countries with better

	Female Labor Force Participation, 1994-2011		
	All Women	Married Women	
	Ancestry	Woman's Ancestry	Husband's Ancestry
Plough Use	-0.062*** (0.020)	-0.136** (0.043)	-0.058** (0.024)
Observations	32,776	6,835	23,124
Adjusted R-sq	0.25	0.11	0.08

Table 8.3. Plough Use and Determinants of Labor Force Participation for Daughters of Immigrants to the United States. Source: Alesina, Giuliano and Nunn (2013, Table IX). This table only shows data for individuals with both parents from the same country. Ancestry= country of origin for parents, and plough use is for that country,. See original paper for other specifications. All regressions include: state-of-residence fixed effects, individual controls, historical country controls and contemporaneous country controls. Column 3 also includes husband controls. See paper for details. *** = significance at 1%, ** = significance at 5%.

gender norms today could have adopted the plough. That would generate a force against the Boserup hypothesis, and controlling for it — as the IV effectively does — would raise our estimates.

That ancestral plough use — no, that the mere *suitability* of geography and climate for crops that facilitated ancestral plough use — should correlate so strongly with outcomes *today*, such as female entrepreneurship or female political participation, speaks eloquently, almost spookily, of the persistence of history. As Mark Twain (apparently) said: history doesn't repeat itself, but it often rhymes.

There is an interesting postscript to the Boserup story as retold by AGN. In general, reduced-form models such as these are good at picking up correlations but less helpful in enlightening us just why those correlations are present. In the present context, one might wonder about the channels of influence running from ancestral plough use to gender norms today. Two obvious pathways come to mind. One, of the Sokoloff-Engerman variety, would argue that informal and formal institutions — such as legal inheritance rights or formal restrictions on participation in politics and business — were affected by ancestral attitudes, which allows history to find its rhyming echo in modern gender-based outcomes. Another pathway would stress attitudes and culture that persisted over time even when their original impulses rooted in agrarian production had died out. AGN make some headway on this question by looking at *attitudes* today and not just outcomes, and indeed there is support for the second pathway, which complements the institutional channels that we emphasized earlier.

A particularly fascinating demonstration of this is in Table 8.3. Consider the daughters of *immigrants* into the United States or Europe, and code ancestral plough use for each such daughter by looking at the country of origin of their parents, and using the plough use index attached to that country as constructed earlier. Pause for a moment to think about the daughter — she is so many steps removed from the plough. She has grown up in the United States or Europe. Her parents' country is part of her inherited "social DNA," but it is not a country she lives in, or possibly even has had much contact with. And yet, even after controlling for all sorts of characteristics, such as marital status (and the socioeconomic characteristics of her spouse if married),

the ancestral plough use index *affects her labor-force participation in the United States today*. Long shadows indeed.

Persecution Perpetuated

^a Hatred and discrimination are norms too, and they can last for millennia. Consider the large-scale anti-Semitic killings during 1348–50 in the wake of the Black Death, for which Jews were blamed. They link the pattern of these killings to the infamous genocide of Jews perpetrated by Germany during the Third Reich. Defeat during World War I had rekindled the flames of anti-Semitism in Germany, leading to persecution on a wide scale in the 1920s and 1930s. But here’s the interesting thing: the *degree* of anti-Semitism varied across locations both in medieval times and in the interwar period. Voigtländer and Voth (2012) explore this variation to examine the persistence of hatred. They show that towns and cities that persecuted their Jews in the Middle Ages were also more likely to display anti-Semitism in the 1920s and 1930s.

When the Black Death struck in Europe in 1348–50, it caused a demographic catastrophe without precedent. As populations searched for an explanation for this sudden epidemic, their attention turned to the Jews. After one tortured Jew “confessed” to poisoning the wells, pogroms occurred in many Northern European towns. There were attacks in Switzerland, Northern France, Germany, and the Low Countries, often before the Black Death reached them (Cohn 2007). At the same time, not all towns with Jewish communities witnessed attacks. Voigtländer and Voth observe that out of 293 towns with confirmed Jewish settlements in the 14th century, 214 recorded pogroms (73%), while 79 did not. In some towns the city authorities or the local prince opposed the burning of the Jews. When local feelings ran particularly high, however, these efforts failed (Foa 2000).

		1349		
		No	Yes	Total
1920s	No	78	196	274
	Yes	1	18	19
	Total	79	214	293

Table 8.4. Pogroms in 1349 and the 1920s. Source: Voigtländer and Voth (2012).

Now fast forward to the tumultuous early years of the Weimar Republic. Several anti-Semitic parties were founded; some achieved a certain degree of electoral success. There were waves of hate speeches, desecrations of graveyards, and violent attacks on Jews. In some resorts, as early as in the 1920s, Jews were declared “unwelcome.” Table 8.4 shows how the frequency of pogroms stacks up in the 1920s, depending on whether the locality had pogroms in the 14th century. Of the 19 pogroms recorded, fully 18 occurred in towns and cities with a record of medieval violence against Jews. One way to read this table is to note that the chances of anti-semitic violence went up from 1/79 (1.3%) in locations without 14th century attacks to 18/214 (8.4%), an increase by a factor of six. The performance of the Nazi Party at the polls in 1928 points in the same direction. Before 1930, the party campaigned on a platform of extreme anti-Semitism. Thus, pre-1930 votes for the Nazis are a particularly suitable indicator for local anti-Semitic sentiment (Heilbronner 2004). Indeed, the Party received 1.5 times as many votes in places that had burned Jews in 1348–50 as in those that had not.

Similar effects exist for the number of Jews deported after 1939, and for the widespread attacks on synagogues in 1938, during the so-called Night of Broken Glass. Wherever Jews were attacked in 1350, Germans in the 1920s and 1930 were more willing to engage in anti-Semitic acts. The same pattern holds for anti-Semitic letters by city dwellers published in the inflammatory Nazi pamphlet *Der Stürmer*.

In summary, there is a consistent pattern of association across a variety of indicators of anti-Semitism between attitudes and actions in the 1920s and 1930s on the one hand, and medieval pogroms on the other. Studying the propagation of such norms and cultures is fascinating and important. The specific case of anti-Semitism may well have formed part of a broader pattern of beliefs about the role of outsiders vs. insiders: an attitude propagated in parental teachings over generations.

^aI am immensely grateful to Nico Voigtländer and Joachim Voth, who sent me the notes on which this box is based. Those notes are based, in turn, on their important research; see Voigtländer and Voth (2012).

8.8. Summary

This chapter has studied how History can cast a long and persistent shadow over subsequent economic development. We studied a variety of different situations from both a theoretical and an empirical perspective. The resemblance to the previous chapter is strong. There, different expectations can lock the same system into multiple equilibria. Here, different histories can lock the same system into distinct steady states.

There are many studies that empirically test for the long shadow of history. The state variables of the past — capital stocks, historical savings rates or fertility, the type of colonial settlement, early modes of production, norms and attitudes, can continue to influence modern socio-economic outcomes. This is viewpoint deeply founded in the negation of convergence.

The examination of such hypotheses in a convincing way isn't easy, even if it is possible to measure these early variables and successfully correlate them with present-day outcomes. But with a mix of creative technique and plain common sense, progress is possible. I've described several studies that attempt to do so. I chose them because they each capture, in a particularly compelling way, the persistent echoes of long-ago history. There are many others. [To be included: references to Banerjee and Iyer, Hoff and Pande, Michaelapoulos, Sascha Becker and Gutenberg, another Voigtlander and Voth study on the horsemen, Alwyn Young on the AIDS epidemic ...](#)

None of this is to argue that history invariably pins down everything. We have discussed both theoretically and with empirical examples how a society can sometimes break free of history, perhaps not always with salubrious consequences. The example of dam building is one such instance. The Arab Spring example in Chapter 9 is another, as are revolutionary movements more generally. The study by Munshi and Myaux (2001) of fertility transition in Bangladesh represents another kind of example altogether. In the context of religious change, the Protestant reformation is a prime example. History is broken when there are great industrials or scientific discoveries. Immediate historical calamities can also be shrugged off when even longer histories have generated aspirations for a better life: think of the rebuilding of Germany or Japan after the disaster of the Second World War.

History casts a long shadow, it is true. We need to be aware of its echoes within us even as we live our daily lives. Theories of convergence that negate history altogether are both impoverished and misguided. But at the same time, that does not entirely negate the possibility of change: the possibility that an entire society can shrug off its past — at least to some degree — and do a lot better, or perhaps a lot worse.