2. The coevolution of institutions and preferences: history and theory*

Samuel Bowles

2.1. INTRODUCTION

Memorable titles, *The Protestant Ethic and the Spirit of Capitalism* (Weber, 1958) or *Nations and Nationalism* (Gellner, 1983) for example, at once capture a historical commonplace and pose a theoretical challenge, namely that the emergence of new institutions is often accompanied by cultural innovation. The appearance of private property in the means of production that was coincident with the domestication of plants and animals 11 thousand years ago and the concomitant eclipse of the egalitarian and collectivist cultures of hunting and gathering bands exhibits this co-evolutionary dynamic. So too does 20th century transformation in family structure and the spread of feminist values.

The theoretical challenges posed by these and similar cultural-institutional dynamics are the following.

First, the processes of institutional and cultural change are frequently the result of a very large number of substantially uncoordinated actions. The novel institutions and cultures that are observed historically are the emergent properties of these decentralized processes, more akin to the evolution of language than to the deliberate social engineering by mechanism design or the dyadic strategic interactions of classical game theory. Leon Trotsky (1970) exaggerated when he wrote that history is 'the natural selection of accidents' but his quip captures an important aspect of human social dynamics that is missing from most economic models.

Second, while entire schools of thought are formed around the purported direction of historical causation connecting culture and institutions – as Weber's and Marx's contrasting views on religion suggest – the examples mentioned above implicate a simultaneous dynamic in which complementary cultures and institutions may jointly emerge and proliferate when neither could survive in isolation.

Third, the very long term persistence of dysfunctional cultural and institutional systems catalogued by Edgerton (1992) casts doubt on the 'efficient design' assumption invoked in biology and by some economists such as Williamson (1985, p. 394): 'viable modes of economic organization ... ordinarily possess an efficiency advantage'.

Finally, the cultural and institutional innovations mentioned above exhibit a striking pattern of long term stasis punctuated by brief periods of rapid change.

To address these challenges I draw on both stochastic evolutionary game theory and biologically-inspired models of cultural evolution (Young, 1998; Boyd and Richerson, 1985; Cavalli-Sforza and Feldman, 1981) to develop a model of the coevolution of cultures (modeled as the distribution of types of preferences in a population) and institutions (the distribution of types of contracts).

Institutions are sometimes modeled as games; the firm, for example is represented by a principal agent model of employer employee interaction. If we are interested in the process of institutional evolution, however, we need to depict institutions not as exogenously given constraints but rather as the *outcome* of individual interactions. In other words, we want to go 'behind' the game describing the institution to investigate the interactions from which it evolved. To do this we specify an underlying game which has as it as its possible outcomes a number of different ways that the participants might interact (Aoki, 2001; Bowles, 2004). The *outcomes of this underlying game* are thus institutions; the process of institutional change will then be studied as a change from one to another of these outcomes.

An insightful way to describe the outcomes of the underlying game is to say that they are conventions, that is, Nash equilibria of an n-person game in which individual adherence to the conventional behavior is a best response as long as the individual believes that a sufficient number of others will also adhere to the convention (Lewis, 1969; Young, 1995). Institutional change occurs when one convention is displaced by another. Thus, institutional innovation and change becomes a problem of *equilibrium selection*. An advantage of this apporoach is that it represents institutional change as the emergent property of the actions taken by very large numbers of people with a characteristic pattern of long term stasis interrupted by brief periods of transition. The very long term persistence of customs governing crop shares and wealth inheritance, their occasional rapid change and their striking pattern of local homogeneity and global heterogeneity are all consistent with this modeling approach (Young and Burke, 2001; Cowlishaw and Mace, 1996).

I model cultural change in a similar decentralized manner. We acquire preferences through genetic inheritance and cultural learning. Because both

are influenced by economic and other institutions, preferences are endogenous. The model I develop highlights the way that structures of social interaction influence the direction and pace of the evolution of preferences. Preferences may be endogenous in other ways. For example, religious or political indoctrination and advertising are undoubtedly important. But the available empirical studies of preferences for brands of food, soap, movies, and other consumption items for which one would expect an important deliberate inculcation effect advertising appears to be less important than one's personal contacts and other influences. Preferences are like accents; we can try to acquire them – learning to love Prokofiev and escargot, or adopting an 'upper class accent' – but for the most part we are only dimly aware of how we acquired them. For this reason the model below is patterned after studies of language change. On the basis of intensive empirical study of linguistic change in Philadelphia, for example, William Labov concluded that

> linguistic traits are not transmitted across group boundaries simply by exposure in the mass media or in schools. ... Our basic language system is not acquired from school teachers or from radio announcers, but from friends and competitors: those who we admire, and those who we have to be good enough to beat (Labov, 1983, p. 23).

The inference is not that institutions such as schools and churches are unimportant, but that we can understand their evolutionary importance may be enhanced by seeing them – along with markets, firms, families, and neighborhoods – as distinct patterns of social interaction affecting the differential replication of behavioral traits rather than simply transmitting the explicit message of the curriculum or the canon Bowles and Gintis (1976).

I begin with four historical cases of cultural-institutional change before attempting a model capable of illuminating their main features.

2.2. CULTURAL AND INSTITUTIONAL DYNAMICS: EXAMPLES

In October 1989, the general secretary of the East German Communist Party, Erich Honecker, grandly celebrated the fortieth anniversary of the founding of the German Democratic Republic as a 'historical necessity' and a 'turning point in the history of the German people'. Parades and demonstrations commemorated the event. Anti-regime protesters had mounted a dozen or so demonstrations over the summer months, but they had attracted fewer than 10,000 participants in all. But twelve days after his address, Honecker resigned, as anti-regime demonstrations mounted first in Leipzig and then

throughout the country, with a million and a half participating in October and twice that number in November (Lohmann, 1994). Within a month, East and West Germans danced on the Berlin Wall, and then dismantled it. Less than a year after Honecker's commemoration of the German Democratic Republic, it passed out of existence, its territory joining the Federal Republic of Germany. As a result, the citizens of the former Communist nation passed from one system of governance to another with an entirely new set of property rights and political processes. Few had anticipated the suddenness and extent of these and the other dramatic changes in institutions which took place throughout most of the formerly Communist world over the same period.

A less heralded but equally dramatic process of institutional and cultural change concerns the painful and dangerous practices of female circumcision and other forms of female genital cutting (FGC) in many parts of Africa (Mackie, 1996, 1999). Like foot binding, which was once widely practiced in China, in parts of Africa FGC is a convention to which families adhere in order to ensure that daughters will be able to marry. The spread of schooling and other modernizing influences on the continent during the 20th century had left FGC intact; in some regions it was spreading. At the turn of the twenty-first century, it was estimated that two million African girls suffered the practice every year.

But at a meeting in 1997 in the small Bambara village of Malicounda in Senegal, residents pledged they would *all* reject FGC. The about face in Malicounda had been prompted not by an anti-FGC campaign but rather by a non-governmental organization which had brought women together to promote literacy and to consider community development and health problems. In nearby Keur Simbara, villagers prudently decided to consult with all of the other villages in the relevant marriage pool; eventually all 13 of these villages collectively pledged to abandon the practice. After village-level meetings, representatives of another cluster of 18 villages of the Fulani ethnic group did the same. Pledge groups spread from village to village. Within a year of the Fulani declaration, the Government of Senegal outlawed FGC.

A third example of institutional and cultural change comes from the Philippines, where the traditional contract governing the rice harvest is called *hunusan* ('sharing' in Tagalog). According the *hunusan* system, any members of the community may participate in the harvest of a farmer's fields, receiving one-sixth of the amount they personally harvest (Hayami , 1998; Hayami and Kikuchi, 1999). The farmer may not rightfully deny anyone this right, and by custom his own family members may not participate, nor may he or they supervise the work. During the 1960s, the one-sixth share provided a return to an hour's harvesting work on a par with

wage-earning opportunities in rural areas, constituting a kind of equilibrium between traditional and modern contracts. However, the increased use of high yielding varieties of rice during the 1970s and 1980s (the Green Revolution) nearly doubled yields and thus greatly increased the value of the one-sixth of the harvest stipulated by the *hunusan* system. As a result, by the end of the 1970s, harvesting under the *hunusan* system earned a substantial rent above the laborer's next best alternative (wage labor).

Some of the large landowners sought to take advantage of the change by reducing the crop share to one-ninth, but this violation of custom caused much indignation among the harvesters, perhaps explaining the suspicious nighttime burning of unharvested crops. The larger farms subsequently invested heavily in both mechanical threshers and in the supervision of harvest labor. Smaller farms, however, continued to offer the one-sixth share, but added to it a traditional obligation that had long been common in some adjacent regions. This was the restriction of harvest labor to those who had performed unpaid weeding services throughout the prior growing season. In contrast to the strategy adopted by the large owners, the new obligation imposed by the smaller farmers did not violate the reciprocity-based *hunusan* system. Hayami (1998, p. 45) reports that 'in the minds of the villagers weeding with no direct payment is considered. ... an expression of gratitude by laborers for the goodwill of the farmer patron who provides them with a guaranteed stable income ...at a time honored share...'. By amalgamating two traditional contracts, the modified *hunusan* system depressed the de facto remuneration of harvest labor to almost exactly the rate for equivalent wage work, thus eliminating the rents introduced by the Green Revolution.

A final example. The labor market aspects of South African *apartheid* were a convention (or a set of conventions) regulating the patterns of racial inequality which had existed throughout most of South Africa's recorded history and had been formalized in the early 20th century and especially in the aftermath of World War II. For whites, the convention was: Offer only low wages for menial work to blacks. For blacks: Offer one's labor at low wages, do not demand more. These actions represented mutual best responses: As long as (almost) all white employers adhered to their side of the convention, the black workers' best response was to adhere to their aspect of the convention, and conversely. Apartheid can be described as a convention because other non-racial and more egalitarian mutual best responses, were in principle feasible even if inaccessible by means of individual action.

The power of apartheid labor market conventions is suggested by the fact that real wages of black gold miners did not rise between 1910 and 1970, despite periodic labor shortages on the mines and a many-fold increase in productivity (Wilson, 1972: Wood, 2000). But a series of strikes beginning

in the early 1970's and accelerating after the mid 1980's signaled a rejection of the convention by increasing numbers of black workers. In doing this, of course, workers and others were not conforming to the best responses of the *status quo* apartheid convention. Their deviation from the convention provoked further defections; many business leaders concluded that the old convention was no longer a best response, leading them to alter their labor relations, raising real wages and promoting black workers. Non-best-response actions by black workers had altered the employers' best responses; as a result the convention unraveled. Within a decade the entire system of apartheid had been abandoned and an ethic of multi-racialism proliferated among whites.

As in the case of the abandonment of FGC in Senegal, the official end of Apartheid – the release of Nelson Mandela from prison and the first all race elections – was the consequence, not the cause of the defacto cultural and institutional rejection of racism by businesses and workers.

2.3. STATIONARY INSTITUTIONS AND PREFERENCES

The acquisition of new preferences may be modeled as a cultural evolution process in which individuals periodically update their behavioral norms (perhaps frequently, perhaps only during adolescence) after having taken into account information about the frequency distribution of various behavior in the population, the payoffs associated with various behaviors in recent periods, and other facts (Bowles, 1998). Equilibrium (that is, stationary) preferences will depend on the nature of the updating rules and the structure of social interactions given by the society's institutions. The latter are important as they determine who meets who to do what tasks and with what benefits. Among the institutions making up this cultural environment are the structure of markets, contracts, legislation and other aspects of society affected by public policy. A consequence is that differing economic institutions support different equilibrium preferences.

This cultural evolution model provides a dynamic setting for the processes studied by Brown et al. (2004). They designed a market experiment to explore the effects of contractual incompleteness on the pattern of trading. The good to be exchanged varied in quality, with higher quality more costly to provide. In the complete contracting condition, the level of quality promised by the supplier was enforced by the experimenter, while in the incomplete contracting condition the supplier could provide any level of quality (irrespective of any promise or agreement with the buyer). Buyers and sellers knew the identification numbers of those they were interacting with, so they could use information they had acquired in previous rounds as a

guide to whom they would like to interact with, the prices and quality to offer, and the like. Buyers had the opportunity to make a private offer (rather than broadcasting a public offer) to the same seller in the next period, thus attempting to initiate an on-going relationship with the seller.

Very different patterns of trading emerged under the complete and incomplete contracting conditions. In the first, 90 percent of the trading relationships lasted less than three periods (most of them were single-shot). By contrast, only 40 percent of the relationships were this brief under the incomplete contracting condition and most traders formed trusting relationships with their partners. Buyers in the incomplete contracting condition offered prices considerably in excess the supplier's cost of providing quality. When Buyers were disappointed by the quality supplied, they terminated the relationship, thereby withdrawing the implied rent from the supplier. The differences in behavior under the two treatments were particularly pronounced in later rounds of the game, suggesting that the traders learned from their experiences, and updated their behaviors accordingly.

These results suggest that trust and reciprocity may depend on the form of the contract, contractual incompleteness sometimes supporting trusting and reciprocal behaviors. The converse is also true: expectations of lower levels of trust and reciprocity would plausibly lead those designing contracts to be willing to pay more for more complete contracts. The implication is that changes in the legal and policy environment that more closely approximate the ideal of complete contracting may have deleterious effects due to their influence on preferences. Behavioral experiments and evidence from natural settings suggests that because explicit economic incentives and other regarding preferences are often sub-modular (substitutes rather than complements) these 'crowding out' effects may be pervasive and substantial (Bewley, 1999; Frey and Jegen, 2001, Bowles, 2008; Bowles and Polania Reyes, 2009).

Based on this evidence of non-separability of other regarding preferences and economic incentives I model an interaction similar to that in Brown et al. (2004) embedded in a dynamic cultural evolutionary environment. As in Bohnet et al. (2001), the underlying process jointly determines the distribution of contracts and the distribution of behavioral norms in the population, a dynamic sometimes termed the *co-evolution of institutions and preferences*

Consider a population of buyer and sellers who are paired randomly for a single interaction. They trade a good or service whose quality (high (H) or low (L)) is determined by the seller and is costly for the buyer to determine *ex ante*. Buyers offer a contract, following which sellers determine the quality of the good they will provide. The buyer may offer one of two

contracts. If the complete (C) contract is offered, the seller receives a fixed compensation just sufficient to offset the costs of providing low quality. These are C-type buyers. According to the incomplete (I) contract, the buyer pays the cost of producing low quality, plus half of the net profits resulting from the transaction. These are I-type buyers.

Sellers are also of two types. R-type sellers interpret the I-contract as a sign of trust on the part of the buyer, and reciprocate by providing high quality, incurring an additional cost of δ_H as a result. When offered a C-contract, however, R-type sellers feel mistrusted, experiencing a subjective cost δ_C, and they retaliate, providing low quality. S-type sellers are completely self-regarding and provide low quality irrespective of the contract. The net surplus of the transaction (net of compensating the seller sufficient to offset the cost of low quality) are π^H and π^L for high and low quality respectively. Those offering a C-contract must pay a cost for monitoring and contractual enforcement, and this cost is greater when interacting with an (offended) R-type (μ_R) than with an S-type (μ_S) with $\mu_R > \mu_S$. Those offering an I-contract make themselves vulnerable to an expected loss of κ through theft should they interact with a S-type seller who gains an additional amount κ as a result.

To exclude cases in which only one of the pairs of contracts and preferences is a Nash equilibrium, I further assume that $\pi^H > 2(\pi^L - \mu_R)$, $\pi^H - \pi^L > 2(\kappa - \delta_H)$ and $\pi^L > 2(\kappa - \mu_S)$. The payoffs appear in Table 2.1. From these assumptions we know that {I,R}, that is, the I-contract matched with the R-seller is the joint surplus maximizing outcome. Thus buyers will offer I contracts if there are sufficiently many R-sellers in the population, and sellers will remain reciprocal if there are sufficiently many I-buyers in the population. But that does not guarantee that {I,R} will be observed in practice in a dynamic setting.

Table 2.1. Payoffs Row player's first. Those in bold refer to the two pure strategy Nash equilibria of this game.

Buyer ↓ \ Seller →	Reciprocator (*R*)	Selfish (*S*)
Incomplete contract (I)	$\boldsymbol{\pi^H/2, \pi^H/2 - \delta_H}$	$\pi^L/2 - \kappa, \pi^L/2 + \kappa$
Complete contract (C)	$\pi^L - \mu_S - \delta_C$	$\boldsymbol{\pi^L - \mu_A, 0}$

2.4. THE DYNAMICS OF PREFERENCES AND INSTITUTIONS

At the beginning of each period with a given probability each member of the sellers and buyers sub-populations may update their strategy upon exposure to a 'model' (a competitor, a teacher, a co-worker, or a neighbor, for example) drawn respectively from the sub-population of seller and buyers, respectively. If the model and the individual have the same strategy, it is retained by the individual. But if the individual and the model have different strategies then the individual may switch if the other strategy would have gained higher payoffs given the distribution of play in the previous period, the probablity of a switch occuring increasing in the difference in the payoffs of the two strategies.

Writing the fraction of the sellers who are reciprocators as ω, the expected payoffs to buyers offering the I- and C-contracts are:

$$v^I = \frac{\omega \pi^H}{2} + (1-\omega)\frac{\pi^L}{2-\kappa}$$

$$v^C = \omega(\pi^L - \mu_R) + (1-\omega)(\pi^L - \mu_S) = \pi^L - \omega\mu_R - (1-\omega)\mu_S$$

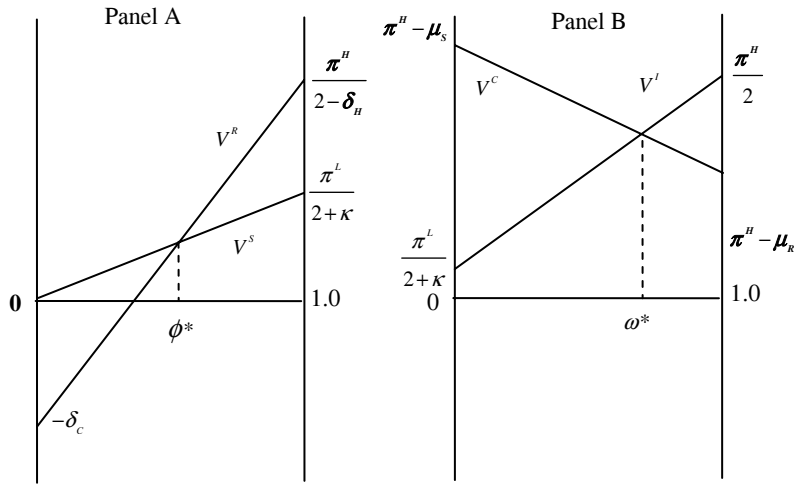
Similarly, writing the fraction of the buyers offering incomplete contracts as φ, the expected payoffs to the R- and S-sellers are

$$v^R = \omega\frac{\pi^H}{2-\delta_H} + (1-\omega) - \delta_C$$

$$v^S = \omega\frac{\pi^L}{2+\kappa}$$

These payoff functions are illustrated in Figure 2.1. What kinds of contracts and behaviors would we expect to observe in this population? One's intuition is that likely outcomes would include a high frequency of both incomplete contracts and reciprocating sellers or the opposite: a predominance of both complete contracts and self-interested sellers. These correct intuitions are readily formalized.

The dynamical system we want to study concerns the state space defined by all possible combinations of contractual and behavioral strategies or $\varphi \in [0,1]$ and $\omega \in [0,1]$. We wish to explore the movement of both φ and ω over time. Letting β be a positive constant that converts payoff differences into the probability that an individual will switch strategies, it is easily shown (Bowles, 2004) that this process gives the familiar replicator dynamic equations



Note: φ is the fraction of buyers offering incomplete contracts while ω is the fraction of sellers who are reciprocal.

Figure 2.1. Expected payoffs to Reciprocal and Self-regarding Preferences (left panel) and Complete and Incomplete contracts.

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \varphi(1-\varphi)\beta(v^I - v^C)$$

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \omega(1-\omega)\beta(v^R - v^S)$$

The terms $\varphi(1-\varphi)$ and $\omega(1-\omega)$ measure (for the buyers and seller respectively) the expected frequency of pairings with 'unlike' models, while $\beta(v^I - v^C)$ and $\beta(v^R - v^S)$ capture the payoff monotone probability of a switch when unlikes meet.

The stationary values of φ and ω in this dynamic are:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = 0 \text{ for } \varphi = 0, \varphi = 1 \text{ and } \omega = \omega^* = \frac{\pi^L + 2(\kappa - \mu_S)}{\pi^H - \pi^L + 2(\kappa + \mu_R - \mu_S)}$$

and

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = 0 \text{ for } \omega = 0, \omega = 1 \text{ and } \varphi = \varphi^* = \frac{2\delta}{\pi^H - \pi^L - 2\kappa}$$

The resulting dynamical system is illustrated in Figure 2.2, with ω and φ the fractions of R-sellers and I-buyers in the previous period and the arrows indicating the out of equilibrium adjustment given by the equations immediately above. The state (φ^*, ω^*) is stationary but it is a saddle: perturbations away from φ^* or ω^* are not self-correcting. The asymptotically stable states are CS (that is, $\omega^* = 0 = \varphi^*$) and IR ($\omega^* = 1 = \varphi^*$), confirming the above intuition.

In this deterministic setting, the initial state determines which of these two asymptotically stable states occurs, and once at either of these states the population remains there (they are absorbing). But historically such 'lock-in' outcomes are not forever.

To take account of the possibility of escape from 'lock-in' a more realistic dynamic would include stochastic influences on payoffs and occasional idiosyncratic updating of preferences or contracts (that is, acquiring lower rather than higher payoff preferences or contracts through error or experimentaiton.) In a plausible version of this process (e.g. Young, 1998), transitions between IR and CS will occur, but the system will spend more (and depending on the technical details, possibly virtually all) of the time in the one with the larger basin of attraction.

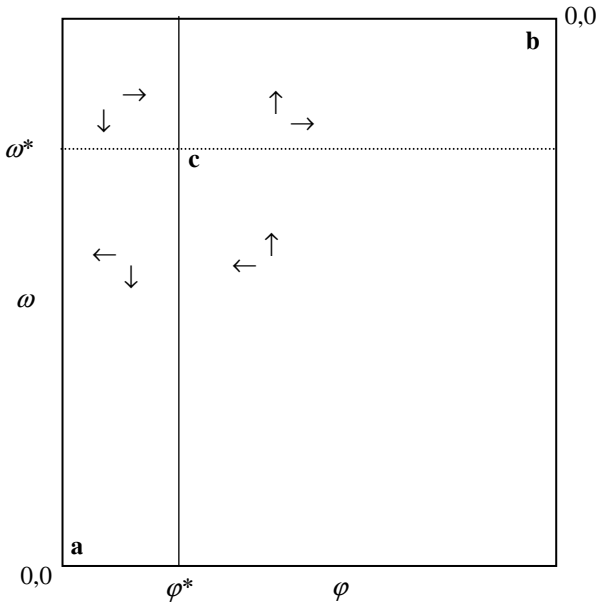


Figure 2.2. The coevolution of the distribution of contracts(φ) and preferences (ω)

2.5. CLASS CONFLICT, TECHNICAL CHANGE AND CULTURAL-INSTITUTIONAL DYNAMICS

Deliberate idiosyncratic may take the form of refusal of the terms of the status quo contract or collective changes in preferences under as a result of religous or political movements. Like accidents or experimentation, these non-best responses may also account for shifts among the two culture-institutional conventions. Suppose for concreteness that the sellers are workers selling their labor time (the effort they provide may be of high or low quality), while those offering contracts are their employers. If the payoffs are as depicted in Figure 2.1 the two classes have different preferences over the two conventions, workers preferring IR while their employers prefer CS. At the CS equilibrium ($\omega^* = 0 = \varphi^*$) employers capture the entire surplus produced when low quality effort is provided by sellers. At the IR equilibrium the larger surplus is shared between the two classes. If for reasons external to the model a fraction of workers greater than ω^* were to become reciprocal, however, the best-responding employers would abandon their preferred arrangements, offering incomplete contracts and thereby inducing best-responding workers also to reject self-interest in favor of reciprocity. Thus ω^* can be interpreted as a kind of collective action threshold necessary for workers to induce a transition to their preferred convention. This framework captures much of the dynamic leading to the demise of apartheid in South Africa, mentioned above.

Suppose now that technical change – a shift from manufacturing to knowledge-intensive production for example – takes the form of increasing the productivity of high quality labor. Because $d\omega^*/d\pi^H < 0$, the consequence is a reduction in the collective action barrier to be surmounted by workers in order to induce a transition to their preferred convention. An increase in μ_S or μ_S associated with increased difficulty in securing even low quality labor through monitoring would (as Figure 2.1 shows) have the same effect. The model thus provides a framework for the analysis of Marx's insight that the dynamism of knowledge and technology may create conditions under which collective action by a class that stands to benefit may overthrow inertial institutions that are unable to fully implement the potential gains associated with technological change.

2.6. CONCLUSION

The model proposed here addresses – albeit in a highly abstract manner – the four theoretical challenges mentioned at the outset: it presents a decentralized process whereby culture and institutions co-evolve, each providing

conditions affecting the dynamic of the other, and in which inefficient institutions may persist over very long periods only to be abandoned as a result of a rapid cultural-institutional transition. The model also provides a framework for analyzing what might be termed quasi-decentralized historical dynamics, namely those involving minimally coordinated collective action by members of the same class or other groups.

Public policy may also affect the long run outcome by altering the two critical values φ^* and ω^*. Consider two possibilities. First, the government imposes the rule of law, reducing theft and setting $\kappa = 0$. Because $d\varphi^*/d\kappa > 0$ and $d\omega^*/d\kappa > 0$ the effect will be to lower the critical fractions of I-types and R-types required to propel the population the surplus maximizing outcome. This is a case of crowding in: an institutional innovation (the rule of law) generates a cultural environment in which reciprocal preferences and hence efficient incomplete contracts may proliferate.

Suppose, as a second example also beginning with the CS status quo, that judicial reforms designed to perfect property rights reduced the costs of contract enforcement, lowering μ_S and μ_R. The effect is to reduce ω^* and to leave φ^* unaffected, so this improvement in property rights enforcement environment increases the critical fraction of suppliers who are R-types necessary to propel dynamic to the efficient equilibrium. In this case crowding out has occurred: a contractual improvement has made an efficient outcome more difficult to attain by raising the barrier to the cultural transition that it requires.

This case illustrates a cultural-institutional analogue to the second best theorem of Lipsey and Lancaster (1956–1957) that may be stated as follows: where contracts are incomplete (and hence socially beneficial values may be important in attenuating market failures), public policies and legal practices designed to more 'improve property rights' so as to more closely align self-regarding preferences and public objectives may exacerbate the underlying market failure (by undermining social values such as trust or reciprocity) and may result in a less efficient equilibrium allocation (Bowles and Hwang, 2008).

A similar model has been used in Belloc and Bowles (2009) to study the effects of international economic integration on the evolution of institutions and culture. We show that two otherwise identical countries will specialize and benefit from gains to trade if their cultural-institutional conventions differ (one at the CS and the other at the IR convention). We also find that the economic integration does not induce convergence to the superior IR convention even when it is Pareto-efficient, but instead renders the institutional-cultural transitions that would accomplish this more costly and hence less likely. Finally, if as above, the sellers are workers and the buyers their employers, in a country at an inferior CS convention, the employers

(but not workers) benefit from trade integration (by contrast to autarchy) while workers benefit from a transition to the 'superior' institutions (but not necessarily their employers).

There are two major lacunae. Both are easily rectified, but not in the short space of a single chapter. First, I have abstracted from competition between territorially based groups such as nations and ethno-linguistic units. This dynamic, essential (if sometimes only implicit) in the contributions of many evolutionary thinkers (Parson, 1964; Hayek, 1988) may be captured by embedding the model presented here in a standard model of equilibrium selection by warfare or other kinds of between-group competition (Bowles, 2006; Choi and Bowles, 2007; Bowles, 2009). If, as is plausible, the probability of success in intergroup contests is increasing in the average payoffs of the group this group selection dynamic will favor the selection of conventions that maximize the joint surplus, providing a second (conflictual but non-revolutionary) mechanism by which Marx's dynamic of institutional change might come about. The model (with or without the group selection addendum) thus provides a mechanism that could implement a kind of selection of efficient institutions, albeit one brought about with long lags by class and inter group conflict rather than individual accommodation to new technological possibilities.

The second shortcoming is that other than specifying a collective action barrier, the model does not explain how collective action – rather than stochastic experimentation or error – could be mechanism by which transitions may come about. This can be done by representing non-best response play as a public good among those members who may benefit from a transition should it occur (for example a strike against employers offering complete contracts). The resulting joint model of the stability of conventions and the process of collective action generating non-best response play (Bowles, 2004) illuminates conditions under which risk dominance or other plausible individual behavioral rules would allow for transitions.

NOTE

* Thanks to Larry Blume, Robert Boyd, Herbert Gintis, Peyton Young and the other participants in the past decade's meetings of the Santa Fe Institute's seminar on the 'Coevolution of Preferences and Institutions' and to the Behavioral Science Program of the Santa Fe Institute, the University Siena, the U.S. National Science Foundation, for support of this project.

REFERENCES

Bewley, T.F. (1999), *Why Wages Don't Fall During a Recession*, Cambridge: Harvard University Press.

Bohnet, I., B. Frey and S. Huck (2001), 'More Order with Less Law: On Contractual Enforcement, Trust, and Crowding', *American Political Science Review*, **95**(1), 131–44.

Bowles, S. (1998), 'Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions', *Journal of Economic Literature*, **36**(1), 75–111.

Bowles, S. (2004), *Microeconomics: Behavior, Institutions, and Evolution*, Princeton: Princeton University Press.

Bowles, S. (2006), 'Group Competition, Reproductive Leveling and the Evolution of Human Altruism', *Science*, **314**(5805), 1569–72.

Bowles, S. (2008), 'Policies Designed for Self Interested Citizens May Undermine "the Moral Sentiments": Evidence from Experiments', *Science*, **320**(5883), 1605–9.

Bowles, S. (2009), 'Did Warfare among Ancestral Hunter-Gatherer Groups Affect the Evolution of Human Social Behaviors?', *Science*, in press.

Bowles, S. and S. Polania Reyes (2009), Economic Incentives and Social Preferences: A Preference-based Lucas Critique of Public Policy, Santa Fe Institute,

Brown, M., A. Falk and E. Fehr (2004), 'Relational Contracts and the Nature of Market Interactions', *Econometrica*, **72**(3), 747–80.

Choi, J.-K. and S. Bowles (2007), 'The Coevolution of Parochial Altruism and War', *Science*, **318**(5850), 636–40.

Cowlishaw, G. and R. Mace (1996), 'Cross Cultural Patterns of Marriage and Inheritance', *Ethology and Sociobiology*, **17**(2), 87–97.

Edgerton, R.B. (1992), *Sick Societies: Challenging the Myth of Primitive Harmony*, New York: The Free Press.

Frey, B. and R. Jegen (2001), 'Motivation Crowding Theory: A Survey of Empirical Evidence', *Journal of Economic Surveys*, **15**(5), 589–611.

Gellner, E. (1983), *Nations and Nationalism*, Ithaca: Cornell University Press.

Hayami, Y. (1998), 'Norms and Rationality in the Evolution of Economic Systems: A View From Asian Villages', *The Japanese Economic Review*, **49**(1), 36–53.

Hayami, Y. and M. Kikuchi (1999), 'Technology, Market, and Community in Contract Choice: Rice Harvesting in the Phillipines', *Economic Development and Cultural Change*, **47**(2), 371–86.

Hayek, F.A. (1988), *The Fatal Conceit: The Errors of Socialism*, Chicago: University of Chicago Press.

Labov, W. (1983), 'De Facto Segregation of Black and White Vernaculars', in D. Sankoff (ed.), *Diversity and diachrony*, Amsterdam: John Benjamins, pp. 1–24.

Lewis, D.K. (1969), *Convention: A Philosophical Study*, Cambridge: Harvard University Press.

Lohmann, S. (1994), 'Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig', *World Politics*, **47**(1), 42–101.

Mackie, G. (1996), 'Ending Footbinding and Infibulation: A Convention Account', *American Sociological Review*, **61**(6), 999–1017.

Mackie, G. (1999), *Female Genital Cutting: The Beginning of the End*, Oxford: St. John's College.
Parsons, T. (1964), 'Evolutionary Universals in Society', *American Sociological Review*, **29**(3), 339–57.
Weber, M. (1958), *The Protestant Ethic and the Spirit of Capitalism*. New York: Charles Scribner's Sons.
Williamson, O.E. (1985), *The Economic Institutions of Capitalism*, New York: Free Press.
Wilson, F. (1972), *Labour in South African Goldmines*, Cambridge: Cambridge University Press.
Wood, E.J. (2000), *Forging Democracy From Below: Insurgent Transitions in South Africa and El Salvador*, Cambridge: Cambridge University Press.
Young, P. (1995), 'Economics of Conventions.', *Journal of Economic Perspectives*, **10**(2), 105–22.
Young, H.P. (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Princeton, NJ: Princeton University Press.
Young, H.P. and M. Burke (2001), 'Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture', *American Economic Review*, **91**(3), 559–73.