The evolution of honesty

E. Somanathan [a,*], Paul H. Rubin [b]

[a] *Planning Unit, Indian Statistical Institute, 7 Shaheed Jeet Singh Marg, New Delhi 110016, India*
[b] *Department of Economics, Emory University, Atlanta, GA 30322-2240, USA*

Received 1 August 2000; accepted 31 October 2002

Abstract

A model of the cultural co-evolution of honesty and capital is analyzed. It is shown that the sign of the payoff differential between honest and dishonest types depends on the ratio of benefits that an employee gets from shirking to the resulting loss of revenue to the firm. If this ratio decreases with capital accumulation, then multiple equilibria in output and honesty are possible in the long run. Small changes in government corruptibility may have large long-run effects on per capita output and the extent of honesty. The honesty and human capital of workers will be positively correlated.

JEL classification: Z13; O10

Keywords: Honesty; Cultural evolution; Social capital; Growth

1. Introduction

This paper analyses the determinants of honesty in market societies, the focus being on the employment relation and the behavior of workers. The conventional economic model assumes that people are honest only insofar as it is in their material self-interest to be so. This is the approach taken by principal–agent theory and by most analyses of problems of opportunism. This narrow definition of honesty not only fails to capture the everyday meaning of the word, but it also ignores conclusive evidence that unselfish honesty is economically significant. Ledyard (1995), in discussing the experimental literature on public goods, shows that there is often more cooperation than economists would predict. When experimental subjects can make (non-binding) promises, they cooperate more than otherwise, implying some level of honesty or unwillingness to lie. Empirical work on taxpayer compliance by Erard and Feinstein (1994) shows that honesty of the disinterested sort is

economically important. Fehr et al. (1997) experiments concerning labor and other markets with moral hazard show that a substantial fraction (35 percent) of subjects supply the effort level stipulated in their contracts even though these are not enforceable. Fehr and Gächter (2000) survey a large literature, both from the laboratory and the field, showing that unenforceable bargains between firms and their employees are frequently concluded and result in wage rigidity.

The level of trust, which is presumably correlated with honesty, appears to vary across societies and is related to a number of important economic, political, and social variables, including the level of per capita income (La Porta et al., 1997; Knack and Keefer, 1997; Fisman and Khanna, 1999). Variations in honest behavior across societies are consistent with the idea that individuals are selfish wealth-maximizers who respond only to incentives, since the variability in honesty could be due to variability in incentive structures. Given the economic importance of "unselfish" honesty, this approach to explaining variations in honest behavior is insufficient. This paper offers a theory of the determinants of unselfish honesty.

We use a biological theory of honesty proposed by Rubin and Somanathan (1998) to explain why individual agents can have preferences to keep their commitments, and why the distribution of people having different preferences will change across generations as incentives change. Evolutionary biology is used to explain how such a flexible distribution of preferences is likely to have evolved.

This flexible distribution of preferences forms the basis of our theory of the determinants of honesty laid out in Section 3. Here, "honesty" refers to something quite specific: a propensity to keep explicit or implicit commitments (to one's employer), and if necessary to forego opportunities to profit at his or her expense. Our central assumption is that the proportion of (unselfishly) honest persons in a society changes slowly in response to the payoff differential between honest and dishonest types. It reaches an equilibrium level when payoffs are equated. We show that the critical parameter that determines the payoff differential between honest and dishonest types, which we call θ, is the ratio of the benefit that a dishonest type gets from shirking to the loss of output caused by shirking.

In general, θ will depend on capital-intensity. In Section 3.2, we examine the interaction between capital accumulation and the extent of honesty in both exogenous and endogenous growth models. If θ is decreasing in the capital-to-labor ratio (as we argue is quite plausible), then the exogenous growth model may exhibit multiple stable steady-state levels of per-worker-output and honesty with higher honesty accompanying higher output. The latter model leads to a unique steady-state growth rate of per-worker-output with an associated unique steady-state level of honesty. Again, if θ is decreasing in the capital-to-labor ratio, higher output is associated with higher honesty. This result is in contrast to those of Bowles (1998a) who finds that greater exposure to markets (which usually accompanies specialization and capital accumulation) reduces the equilibrium proportion of "nice" types, or may drive them to extinction. The key difference between Bowles' assumptions and ours is that he assumes that greater exposure to markets makes it less likely that "nice" players are recognized as such, the rationale being that market transactions are more ephemeral than traditional pre-capitalist ones. Because we consider mainly labor markets in which transactions are, not coincidentally, typically *not* ephemeral, we assume that there is some probability that honest types will be recognized as such. Bowles' theory may be thought

of as an account of the transition from tribal to market society, ours as an account of the development of capitalist society.

In Section 3.3, we show that exogenous decreases in incentives to be honest, such as a weaker and more corruptible government, can lead to a fall in honesty for each level of the capital stock. Such variation can, therefore, lead to the appearance or disappearance of multiple equilibria in output and honesty. Countries with weak and corruptible governments could show negative growth rates and declining levels of honesty as they converge to a low steady state. Section 3.4 examines the within-country association between human capital and honesty. Finally, in Section 4, we discuss possible extensions of the analysis, and conclude.

Before proceeding with the paper, we suggest three reasons why analysis of unselfish honesty is of interest. First, whether some proportion of workers is honest in the everyday meaning of the word (and what that proportion is) has very important implications for the nature of labor markets and the internal organization of firms. The existence of honest types who cannot be distinguished from dishonest ones will, in general, change the incentive scheme that it is optimal for a firm to offer to workers. In fact, the nature of the optimal incentive scheme will depend on the proportion of honest types, h.[1]

To see why, consider the following simple example of an agency problem. There are two possible verifiable revenue levels, bad and good, for the principal/firm, and two possible unverifiable effort levels, low and high for the agent/worker. The probability of the good revenue level conditional on high effort is greater than its probability conditional on low effort. The firm is risk-neutral while the worker is risk-averse. If the worker is dishonest, that is, will shirk unless the incentive scheme pays him not to do so, then the optimal scheme involves paying more when success occurs (provided that is better for the firm than having the worker shirk, which we shall assume). If the worker were known to be honest in the sense of keeping any promise to exert high effort despite the impossibility of detecting shirking, then the optimal contract would involve perfect insurance for the worker. For values of h close to one, the firm does best to provide perfect insurance. As h falls, the revenue loss from shirking by dishonest types increases until h reaches a level where it pays the firm to switch to the contract that induces dishonest types to exert high effort. This point is reached when the welfare loss from forcing workers to bear more risk is outweighed by the gain in expected output that results from the high effort level now exerted by dishonest as well as honest types.

Second, economic theory has traditionally assumed stable preferences, one of the most important reasons for this assumption being that allowing for changing preferences makes it hard to make predictions. This paper illustrates that preferences can vary, but are nonetheless predictable.[2]

Finally, of course, honesty has an *intrinsic*, as well as an instrumental, value and therefore is of interest in itself.[3]

[1] Fehr and Gächter make this point at some length in a different context.

[2] Bowles (1998b) surveys the (largely non-economic) literature on the cultural evolution of preferences.

[3] The distinction between intrinsic and instrumental values and the necessity of going beyond traditional welfare criteria have been emphasized by Sen (1979).

2. Honesty and cheating

The central assumptions underlying our theory about variability in honesty in societies is that people may be socialized to be honest, largely in childhood, and that the extent of this socialization responds to the return to honesty relative to dishonesty. This section explains why evolutionary theory would lead us to predict that parents will socialize their children in this manner and discusses related evidence from psychology and other disciplines.

The basic evolutionary assumption is that humans have those preferences that served to increase genetic fitness in the applicable past, the "Environment of Evolutionary Adaptedness" (Barkow et al., 1992). A large literature (Trivers, 1971; Hirshleifer, 1987; Frank, 1987, 1988; Güth and Kliemt, 1994) has argued that humans have evolved emotional predispositions that enable them to be loyal even when it is not in their material or reproductive interest to be so. Frank (1988) shows that they signal these predispositions through their demeanor. Others, reading these signals, realize that it is safe to enter into cooperative ventures with them, since the likelihood of being cheated is low. This raises the material, hence reproductive, payoff of the possessor of the emotional trait of loyalty. Of course, in this situation, it pays to develop the signals *without* the underlying disposition to be loyal, so that one can cheat and get even higher payoffs. Such mimicry is often seen in the natural world. There is then an evolutionary race between the genuine article and the mimic, with the one evolving to differentiate itself and the other evolving to copy the signals. Both types may end up being present in the population. In these situations, there will also be selection pressure to enable people to detect cheating. Cosmides and Tooby (1992) report experimental results from the Wason Selection Task that suggest that humans are well-adapted to do this.

Frank's model implies that human beings will consist of a mix of people, some who keep their commitments and some who do not. The experimental literature cited above confirms this prediction. Moreover, there is evidence that it is difficult for people to pretend to be honest and that it is possible for people to detect true honesty through interactions with others. Frank (1988) makes this point at length. Frank et al. (1993) have shown experimentally that with only a one-half-hour interaction, subjects' predictions regarding who is likely to defect in a prisoner's dilemma game are substantially better than random. Brosig's (2002) experiments confirm the finding, under more stringent conditions, that subjects' ability to detect honesty in even brief face-to-face interactions with strangers is better than random.

Rubin and Somanathan point out that environmental changes during the Pleistocene would have shifted the payoff functions of honest types and mimics, so that at certain times and places, honest types would have done better than dishonest types while at others, this payoff differential would have been reversed. Therefore, natural selection would favor a strategy capable of switching from honest behavior to dishonest behavior and back as circumstances changed. Such a flexible strategy would then drive genetically hardwired honest and dishonest strategies to extinction.

It may appear that such switches would not be favored by evolution, since the survival value of honesty lies in an ability to *commit* to honest behavior. However, there is good evidence that honest behavior can be inculcated by parents in their children. A person would then grow up to be honest only if they had been taught to do so as a child. The contributions in Kagan and Lamb (1984), by psychologists and anthropologists, and the

work of the psychologist Robert Coles contain extensive evidence for the ability of parents to inculcate varying moral beliefs and behavior in children (Coles, 1997). Mayr (1997, Chapter 12) presents a similar analysis from a biological perspective. Wilson (1993), a political scientist, makes the same point.[4]

How would dishonest parents socialize their children to be honest? By steering their children into environments (Akerlof's "loyalty filters") that will result in interactions with honest role models and peers and by concealing the extent of their own dishonesty. Honest parents in an environment in which dishonesty pays could socialize their children to be dishonest in like fashion. Moreover, while behavior patterns learned in childhood are often long-lasting, they are not always immune to change, and honesty is no exception. Honest parents may learn that dishonesty pays and change their behavior. None of the conclusions of our theory below depend on honesty being lifelong. It need only be lasting over an economically significant time, such as a significant portion of the expected length of an employment contract.

The possibility of socializing children to behave in particular ways confers adaptability to changing circumstances over the time span of a generation while still enabling individuals to commit to honest behavior once they had been so imprinted. In other words, biological evolution may have endowed us all with the capacity to form emotional loyalties and feel guilt, but we need to be schooled about the circumstances in which these feelings will be triggered. Thus, with the evolution of loyalty and trust being partly cultural, we may expect to see variations in h (which now represents the proportion of people who have been *taught* to be honest) across societies. The extent of honesty in a society can thus respond to incentives over the long run, even though, in the short run, honest individuals ignore incentives to cheat.

3. The extent of honesty in a society

3.1. The equilibrium level of honesty

We shall analyze three issues. First, how does the level of capital affect the extent of honesty and how do they interact? Second, how do exogenous variations in the rewards to dishonest behavior, the most important source being government structure, affect honesty in civil society? And third, how is the distribution of human capital related to the distribution of honesty? We have argued that humans have been selected to possess the ability to commit to honesty and that they have also been selected to indoctrinate their children to be honest or not depending on the material payoffs to honesty and dishonesty.[5] This means that the extent of honesty in society can change over time, growing when the payoffs to honesty exceed those to dishonesty, and shrinking when the reverse is true. We now develop a theory

[4] In economics, Akerlof (1983) presents a model in which parents can socialize their children to be honest or not. See also Noe and Rebello (1994) and Grossman and Kim (2000).

[5] As was pointed out earlier, the assumption that all adaptation to changes in payoffs takes place through parental inculcation is in no way essential. All that is necessary for our results is that there is some ability to commit to honest behavior over an economically relevant period, together with some adaptability of behavior over longer time spans.

of honesty for our current environment, in which there are established business firms and other institutions that hire individuals to work for them. The individuals we consider are those who have evolved in ways described above, so that at any given time, some are honest and can commit to honesty, and others cannot.

We focus on the labor market, since the payoffs to being honest or not are probably most importantly affected through labor incomes. There are two types of workers, honest, H, and dishonest, D. Honest workers are potentially more productive than dishonest ones since they are trustworthy and always keep their commitments.[6] Dishonest workers, on the other hand, may shirk and steal. We intend these terms to cover a broad range of behavior. For example, an engineer who designs a computer system for his firm that is far from being the best he could have chosen but that increases the firm's demand for his services is behaving dishonestly by our definition.

A proportion ψ of honest workers emit a signal ($S = 1$) that reveals their type with certainty. By being around some honest people long enough, the employer forms a judgement about their character and correctly concludes that they are trustworthy.[7] The remaining $1 - \psi$ honest workers, however, emit no signal ($S = 0$). Neither do D's. We assume that when a person is socialized in childhood to be honest, he acquires the distribution $G_{\mathrm{H}}(S)$ of the signal S, but not a particular value of the signal. Thus, with probability ψ, a firm learns the type of an honest worker.

The employer's prior probability h that any given worker is honest is assumed to equal the true proportion of honest workers in the population. Therefore, the probability that a given worker is honest conditional on $S = 0$, is

$$\frac{h(1-\psi)}{h(1-\psi)+1-h} = \frac{h(1-\psi)}{1-h\psi}.$$

Let K denote the capital stock, H the number of honest workers, D the number of dishonest workers, and $H + D = L$ the total number of workers. A D lowers output below the level an H would produce because he may shirk or steal. The production function is

$$Y = F(K, H + \lambda D), \tag{1}$$

$$Y = F(K, [h + (1-h)\lambda]L), \tag{2}$$

where $0 < \lambda < 1$ is a constant that reflects the lower productivity of the D's as compared to the H's. In a slight departure from conventional usage, the production function above represents output appropriated by *firms*. *Total* output is greater than this since some of the output lost to firms due to shirking or stealing by D's is appropriated by the D's.

It is assumed that F is increasing in K and $H + \lambda D$, and it has diminishing returns to $H + \lambda D$. In both the exogenous and endogenous growth models, we assume constant returns to scale and, therefore, diminishing returns to capital at the level of the individual firm. In the

[6] We ignore the issue of the honesty of firms and simply assume that firms keep their commitments to workers. This may be because of reputation effects that are more salient for firms than for workers. Or, as emphasized by Fehr et al., it may be necessary to induce honest workers to keep their commitments to firms, since such workers have a taste for reciprocity.

[7] As Frank (1988, p. 12) puts it: "Do you know anyone, not related to you by blood or marriage, who you feel certain would return it [$1000 in an envelope with your name on it] to you if he or she found it?"

endogenous growth model, constant returns to aggregate capital and, therefore, increasing returns to scale will arise as a result of the external effect of the aggregate capital stock on the output of each firm, as in Romer (1986).

The marginal products of H's and D's, respectively (again, from the firm's point of view, that is, not counting the output appropriated by D's) will be denoted by

$$x_H \equiv \frac{\partial F}{\partial H}, \qquad x_D \equiv \lambda \frac{\partial F}{\partial H} = \lambda x_H. \tag{3}$$

Since part (but not all) of the output lost to the firm is captured by the D, his private benefit is

$$\theta(x_H - x_D),$$

where $0 < \lambda < 1$. The ratio θ of (expected) private benefits that an employee gets from shirking to the resulting loss of revenue to the firm, is a key parameter in what follows and we will return to its determinants in Section 3.2.

Firms are competitive, so that all workers are paid their expected marginal product. Thus, the expected productivity of a worker with signal $S = 0$ is

$$\pi_0 \equiv \left(\frac{h(1-\psi)}{1-h\psi}\right) x_H + \left(\frac{1-h}{1-h\psi}\right) x_D. \tag{4}$$

An H with signal $S = 1$ is paid his marginal productivity x_H.

Therefore, the expected payoff of an H is

$$\pi_H = \psi x_H + (1-\psi)\pi_0. \tag{5}$$

That of a D is

$$\pi_D = \pi_0 + \theta(x_H - x_D). \tag{6}$$

If $\pi_H > \pi_D$, then as parents socialize their children to reflect current returns to honesty, h will rise until either $\pi_H = \pi_D$ or $h = 1$. Similarly, if $\pi_H < \pi_D$, then h will fall until either $\pi_H = \pi_D$ or $h = 0$.

We define an *equilibrium* h^* to be a value of h (given values of K and L) such that $\pi_H = \pi_D$, or $h = 0$ with $\pi_H < \pi_D$, or $h = 1$ with $\pi_H > \pi_D$. Because the payoff functions are continuous in h, an equilibrium exists.

Now by (4)–(6), when $h = 1$, $\pi_H < \pi_D$. So, $h^* < 1$. We now equate payoffs to the two types to see under what conditions an interior equilibrium exists. At an interior equilibrium h^*, with $0 < h^* < 1$,

$$\psi x_H + (1-\psi)\pi_0 = \pi_0 + \theta(x_H - x_D),$$

that is,

$$\psi(x_H - \pi_0) = \theta(x_H - x_D).$$

The left-hand side above is the probability that an H is recognized to be an H times the increase in payoff the H obtains from not being pooled with the D's. That is, it is the expected amount by which an H's payoff exceeds that of an H who is indistinguishable

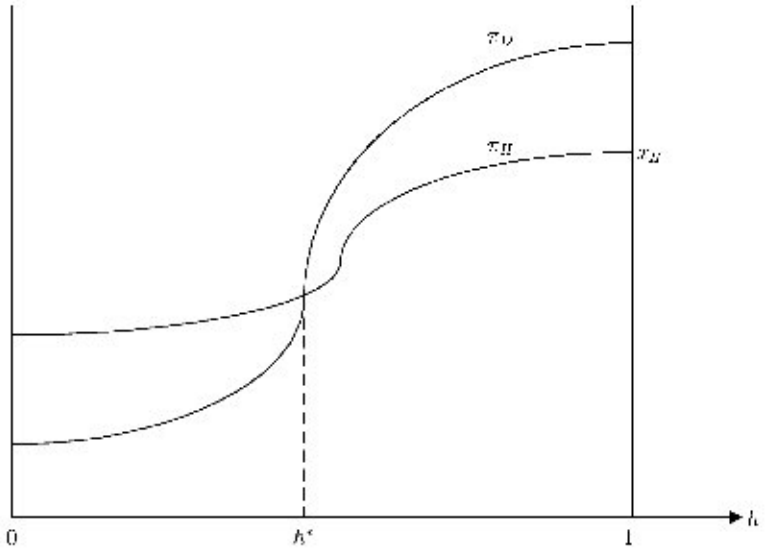

Fig. 1. Payoff functions and an interior equilibrium.

from a D. The right-hand side is the amount by which a D's payoff exceeds that of an H who is indistinguishable from a D.

Now π_0 is a weighted average of x_H and x_D, with the weights depending on h, so we can rewrite the equality above as

$$\psi\left(\frac{1-h}{1-h\psi}\right)(x_H - x_D) - \theta(x_H - x_D) = 0,$$

where the left-hand side is the payoff differential between H's and D's. Thus, the sign (though not the magnitude) of the payoff differential is independent of the level of productivity and depends only on h and the parameters ψ and θ. Now the difference $x_H - x_D$ drops out of the equality above which reduces to

$$h^* = \frac{\psi - \theta}{\psi(1-\theta)}. \tag{7}$$

This is the (unique) equilibrium if $h^* > 0$, which, of course, is the case if and only if

$$\psi > \theta. \tag{8}$$

If (8) does not hold, then the equilibrium is $h^* = 0$. This is not a very interesting case. So, in the rest of the paper, we shall assume that (8) holds. That is, we assume that sufficiently many of the H's succeed in separating themselves from the D's so that the higher payoffs they get outweigh the D's gain from stealing when h is near zero, and, therefore, the equilibrium level of honesty is positive. It is clear that $\pi_H > \pi_D$ for $h < h^*$, and $\pi_H < \pi_D$ for $h > h^*$, so the equilibrium is stable in any evolutionary dynamic (Fig. 1).[8]

[8] A dynamic $\dot{h} = g(h)$ is evolutionary or payoff-monotonic if $\pi_H > \pi_D \Leftrightarrow g(h) > 0$ and $\pi_H = \pi_D \Leftrightarrow g(h) = 0$.

The equilibrium level of honesty, therefore, depends only on the parameters ψ and θ. Since ψ is biologically determined, it is the dependence of h^* on θ, the ratio of benefits from shirking obtained by a D to the loss of output caused by such shirking, which is of interest.

3.2. The level of development and the extent of honesty

We have assumed that honesty has a positive effect on output since it allows for better solutions to problems of opportunism. We have also allowed the extent of honesty to be endogenously determined by payoffs to honesty and dishonesty. In this section, we ask what the feedback effect of income on honesty and trust will be. Given that output may affect honesty, how will output and honesty co-evolve? What relation might we expect to see in the data between trust, per capita income, and related variables? The material in this section will also provide a framework for answers to the questions posed in Section 3.3: what effects will exogenous shifts in honesty have in the long run on honesty and output? Will small exogenous shifts necessarily have small effects?

The basis of our analysis in this section is the observation that θ may be a function of the capital-to-labor ratio $k \equiv K/L$. As the economy gets more capital intensive, the damage that a D does when he shirks is likely to rise since he destroys more capital. While benefits from shirking may rise as well, they may not rise as fast. The benefit of shirking can go up with real income, but the adverse consequences can rise much faster. For example, carelessness at a chemical or nuclear plant or on an oil tanker (as in Bhopal, Chernobyl, and the Exxon Valdez) can lead to vastly greater harm than would be possible in a less capital-intensive economy. Moreover, this fact can lead to increased monitoring of workers not known to be H's, further reducing the benefits of shirking while increased monitoring, due to its costs, contributes indirectly to the reduction of productivity occasioned by the presence of D's (the denominator in θ). Thus, the proportion of the drop in productivity caused by shirking and stealing that is captured by the D may be smaller for larger k. This assumption can be written as

$$\theta'(k) < 0. \tag{9}$$

We will also consider the implications of the alternative, perhaps less plausible, assumption

$$\theta'(k) > 0. \tag{10}$$

From (7), it is clear that $h^{*\prime}(\theta) < 0$, so h^* rises with k if and only if (9) holds.

Under either of the assumptions (9) or (10), the upper and lower bounds of h^* will be approached as k tends to zero or infinity. What these bounds are will depend on the bounds of the function $\theta(k)$. Suppose, for concreteness, that (9) holds. If $\theta(k) > \psi$ for small k, then $h^* = 0$. If the upper bound of $\theta(k)$ is less than ψ, then the proportion of honest workers will remain strictly positive and bounded away from zero even when k tends to zero. Similarly, if the lower bound of $\theta(k)$ is zero, then nearly everyone will be honest when k grows large. But if the lower bound of $\theta(k)$ is positive, then dishonesty will not disappear even if k approaches infinity.

We now examine the long-run behavior of the economy. First, consider an exogenous growth model. We use a Solow model and assume constant savings, depreciation, and

labor-force-growth rates s, δ, and n, respectively, and the Inada conditions $\partial F/\partial K \to \infty$ as $K \to 0$, and $\partial F/\partial K \to 0$ as $K \to \infty$. The production function (2) can be written as

$$y = F(k, (1-\lambda)h + \lambda),$$

where y denotes output per worker. So,

$$\dot{k} = sF(k, (1-\lambda)h + \lambda) - (\delta + n)k. \tag{11}$$

Assume that h is at its equilibrium value h^* in the long run. h^* is a bounded function of k, increasing if (9) holds and decreasing if (10) holds.

Now write

$$\dot{k} = sf(k) - (\delta + n)k \tag{12}$$

where

$$f(k) = F(k, (1-\lambda)h^*(k) + \lambda).$$

So,

$$\frac{\mathrm{d}f}{\mathrm{d}k} = \frac{\partial F}{\partial K} + \frac{\partial F}{\partial L}(1-\lambda)\frac{\partial h^*}{\partial k}. \tag{13}$$

From (12) and the Inada condition $\partial F/\partial K \to \infty$ as $K \to 0$, the growth rate of capital is positive for low levels of k. Since $f(k)$ is bounded above by $F(k, 1)$, the Inada condition $\partial F/\partial K \to 0$ as $K \to \infty$ ensures that the growth rate of capital will be negative for k large enough. Therefore, there exists at least one stable steady-state level of k.

Suppose (9) holds. The increase in honesty that accompanies increases in capital-per-worker leads to a higher steady-state level of capital-per-worker (than would obtain if workers were neoclassically self-interested) and raises the growth rate during the transition. That is, the effect of diminishing returns to capital accumulation is moderated by the accompanying increase in the productivity of the workforce as it becomes more honest. In fact, if h^* increases in k fast enough over some range, it can overwhelm the effects of diminishing returns to capital. This means that output can be convex in k over some intermediate range. Therefore, $sf(k)$ may intersect the line $(\delta + n)k$ more than once at positive levels of k. Multiple stable steady-state levels of honesty and per capita output are, therefore, possible. They will exist provided that $\theta'(k)$ is sufficiently negative in some interval $(k^*, k^* + \varepsilon)$ where k^* is the smallest positive steady-state value of k. This condition will ensure that $f(k)$ rises steeply enough following its first intersection with $(\delta + n)k$ that it intersects it again. An example is graphed in Fig. 2.[9]

The convexity of $f(k)$ in the example is generated by assuming a $\theta(k)$ function that falls slowly at first, then rapidly, finally leveling off as $\theta(k)$ approaches zero, its lower bound. This is certainly conceivable, although it is by no means the only likely case. At low levels of development, increases in the capital stock may not be accompanied by long production chains. At higher levels, increased complementarities may result in much more damage

[9] The parameter values and functional forms used to generate the figure are $f(k, (1-\lambda)h^*(k) + \lambda) = k^{\alpha}[(1-\lambda)h^*(k) + \lambda]^{1-\alpha}$, $\theta(k) = \psi/(1 + 10k^5)$, $\psi = 0.2, \alpha = 0.1, \lambda = 0.3, s = 0.25, n + \delta = 0.26$.

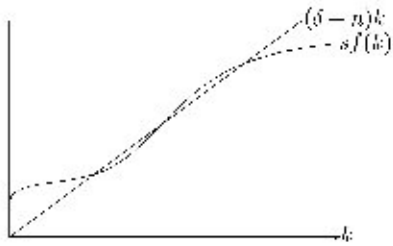

Fig. 2. Exogenous growth with multiple steady states.

when there is shirking and stealing. So, the proportion of output lost to the firm that is appropriated by dishonest workers falls more rapidly. Eventually, as $\theta(k)$ approaches its lower bound, it must once again fall only slowly.

If, on the other hand, (10) holds, then declines in honesty that accompany capital accumulation will slow output growth (see (13)), leading to a lower steady-state level of per capita output than would have been the case if θ did not change with k. In this case, there is a unique steady-state level of k, y, and h.

Next, consider Romer's endogenous growth framework

$$Y_i = K^{1-\alpha} K_i^{\alpha} ((1-\lambda)h + \lambda)^{1-\alpha} L_i^{1-\alpha},$$

where subscripts denote the values of the variables for the ith (competitive) firm and K is the aggregate capital stock. We suppose the total labor force is fixed at unity, so that $K = k$ and $Y = y$. Then,

$$y = k((1-\lambda)h + \lambda)^{1-\alpha},$$

so that

$$\frac{\dot{y}}{y} = \frac{1}{y}[\dot{k}((1-\lambda)h + \lambda)^{1-\alpha} + k(1-\alpha)(1-\lambda)((1-\lambda)h + \lambda)^{-\alpha}\dot{h}]$$
$$\to s((1-\lambda)h^{S} + \lambda)^{1-\alpha} - \delta, \tag{14}$$

as $t \to \infty$.[10] Here, h^S denotes $\lim_{k\to\infty} h^*(k)$.

In this model, since there are constant rather than diminishing returns to capital, the capital-to-labor ratio tends to infinity in the long run. Therefore, the proportion of honest types, which is bounded, tends to its limit. The model then reduces to an Ak model that, like all such models, has a single steady-state growth rate.

If $\theta'(k) < 0$ ((9) holds), then honesty will increase with k. Hence, using (13), we see that as the steady-state growth rate is approached, h^* will approach its upper bound, and the growth rate of output will be rising. On the other hand, if $\theta'(k) > 0$ ((10) holds), then honesty will decrease with k and the growth rate of output will fall towards its steady state.

Evidence that bears on these propositions from the studies by Knack and Keefer and La Porta et al. is mixed. Using data from the World Values Survey, they report a positive

[10] We assume $s > \delta$, so that the growth rate is positive.

correlation between GDP per capita and the proportion of people in a country who say they would trust most people. In addition, Knack and Keefer report a positive correlation between trust and capital per worker. But they also report a decline in trust in the US, the only country for which a time series was available, between the 1950s and 1960s and the 1980s and 1990s. While our measure of honesty and trust would require respondents to report the *proportion* of people who could be trusted, the World Values Survey measure would presumably be correlated with ours. However, the surveys do not control for incentive structures or for the reference group to whom the respondents are referring (most of which people?). So, the surveys will pick up both honest behavior (and therefore trust) generated by incentives to be honest, as well as 'intrinsic' honesty of the kind we are modeling. These findings are consistent with the assumption $\theta'(k) < 0$ (9) in both exogenous and endogenous growth models.[11]

The assumption $\theta'(k) > 0$ (10) is inconsistent with the positive correlation between honesty and per-capita-output reported above (in both growth models), unless we admit the possibility that $h^*(k)$ falls so rapidly with k that y actually declines as k increases. Since this seems contrary to the facts, we can rule it out. However, in Section 3.3, we will qualify this conclusion when we allow for exogenous variability in $h^*(k)$.

3.3. Government corruptibility and honesty

Thus far, we have supposed that the only determinant of the payoff functions for honest and dishonest types is the private-sector labor market. In actuality, these payoffs will be influenced by other factors as well, dependent on the social and political structure of society. For example, ineffective contract and law enforcement by a weak or lawless government can raise the payoffs to dishonest types relative to those of honest types, thus lowering $h^*(k)$. Certainly, one of the important influences on payoff functions is the public-sector labor market. We now discuss its effect on $h^*(k)$ and, through it, on long-run output and honesty.

Differences in the organization of the state can lead to more or less scope for corruption in government. This can result in variations across societies in the payoffs to those H's and D's who are employed in the public sector. A more corruptible state apparatus can raise the payoffs to D's in government relative to H's in government. La Porta et al. find a positive correlation between trust and various measures of government efficiency. Of course, such differences in the corruptibility of the state will be in part endogenous. This is, in fact, La Porta et al.'s interpretation. However, we focus on exogenous differences to isolate their effects.

Suppose there is a public sector that hires a proportion γ of the labor force. If the government is not corrupt, one may suppose that its hiring is identical to that of the private sector so that the model is unchanged. If the government is (sufficiently) corrupt, then we may suppose that it (that is, those who control recruitment for it), will wish to hire dishonest workers. Then, it must set the expected payoff g_D for dishonest workers (wage plus proceeds of theft) to be higher than their expected earnings in the private sector. In order to

[11] It is true that certain small-scale closed societies with very low k exhibit high degrees of honesty and trust. As mentioned in Section 1, however, our theory applies only to capitalist societies.

distinguish dishonest from honest workers who do not signal their honesty, it would also set the payoffs to honest workers, that is, their wages, to be lower than their wages in the private sector.

Let h^*_C denote the equilibrium level of h when the government is corrupt. The assumptions above imply that in this equilibrium all public-sector workers would be dishonest. Dishonest workers would compete for public-sector employment while honest workers would apply only for private-sector jobs. At such an equilibrium, a dishonest worker's chance of getting public-sector employment would be $\gamma/(1-h^*_C)$. The fraction of honest workers in the private-sector labor force will be $h^*_C/(1-\gamma)$. Denoting the payoff function of dishonest workers by $\pi^C_D(h)$, it follows that in equilibrium,

$$\pi^C_D(h^*_C) = \left(\frac{\gamma}{1-h^*_C}\right) g_D + \left(\frac{1-h^*_C-\gamma}{1-h^*_C}\right) \pi_D\left(\frac{h^*_C}{1-\gamma}\right) = \pi_H\left(\frac{h^*_C}{1-\gamma}\right), \qquad (15)$$

unless $h^*_C = 0$, in which case, of course, $\pi^C_D(h^*_C) \geq \pi_H(h^*_C/(1-\gamma))$.

Corruption may reduce the proportion of honest workers to zero. If it does not, then (15) holds. In this case, since $g_D > \pi_D(h^*_C/(1-\gamma))$, it follows that $\pi_H(h^*_C/(1-\gamma)) > \pi_D(h^*_C/(1-\gamma))$. This implies that $h^*_C/(1-\gamma) < h^*$ where h^* denotes the equilibrium when the government is not corrupt and follows hiring practices identical to the private sector. A corrupt government not only has an adverse effect on the equilibrium proportion of honest workers in society as a whole, but also on the proportion of honest workers in the *private* sector.

In the exogenous growth model above, a lower $h^*(k)$ function reduces the growth rate of capital at all levels of k. It also reduces the steady-state level of output-per-worker, if this is unique. Other things equal, more scope for corruption in government will lead a country to be poorer because its workers will be less able to make binding commitments to work well even in the private sector. This labor-market effect is in addition to the usual channels by which governments affect output.

Moreover, if output is convex in k in some range, then a high $h^*(k)$ function could result in the disappearance of multiple steady states that would be generated by a low $h^*(k)$ function. To see this, note that a higher $h^*(k)$ function would mean a higher $sf(k)$ function in Fig. 2. Therefore, even small differences in $h^*(k)$ could lead to large differences in steady-state levels of output-per-worker, not just to different growth rates during the transition to a steady state.

The negative growth rates found in some of the most corrupt countries in the world suggest that multiple steady states may be more than a theoretical possibility. These countries may be converging to a low-level steady state.

Applying shifts in $h^*(k)$ to the endogenous growth model, we see from (14) that a fall in $h^*(k)$ will lower the growth rate of output-per-worker in the long run and not just its level. Thus, *small shifts in honesty can potentially have large effects on output in the long run in both types of growth models*. Note, though, that it is only in the exogenous growth model with multiple steady states that a small shift in $h^*(k)$ can have a large effect on the extent of honesty in the long run. In the endogenous growth case, honesty must tend towards its limiting level in the long run, and a small change in $h^*(k)$ can, therefore, have only a small effect on the long-run level of h.

Now that we have allowed for exogenous shifts in $h^*(k)$, we can no longer firmly rule out the assumption $\theta'(k) > 0$ on the basis of the positive correlation between per capita output and levels of trust reported by La Porta et al. $\theta'(k) > 0$ may hold, yet higher levels of k may still mean higher levels of h and y for exogenous reasons.

3.4. Human capital and honesty

Let the education level or human capital of a worker be denoted by E. Suppose, for simplicity, that output of workers of different human capital levels is additive. Then, the production function for workers with human capital level E can be written

$$Y = F(K, EH + \lambda ED),$$

where we assume that F has constant returns to scale. Note that if E is large, then the marginal product of labor will be high and the wages paid to workers, honest and dishonest, will be high. Therefore, the wage-to-rental ratio will be high, as will capital-per-worker k. Thus, the equilibrium level of honesty for workers with high human capital will be high if h^* is increasing in k. This will be the case if and only if (9) holds, that is, if and only if $\theta'(k) < 0$. Notice, from (7), that h^* depends only on θ and not on the marginal productivities x_H and x_D of honest and dishonest workers. Here, we are assuming that parents would tend to socialize their children to be honest or not according to the payoff to honesty relative to dishonesty, *conditional* on the expected level of human capital their children will accumulate. We are also assuming that their expectations are on average correct.

There will then be a correlation between human capital levels of workers and their honesty, positive if $\theta'(k) < 0$, and negative if the reverse inequality holds.[12]

If capital markets are imperfect, or if acquired human capital is complementary with human capital provided at home, or if productive human capital is partially innate and levels are inherited, then poorer families will accumulate less human capital. To the extent that capital markets are imperfect and wealth is a determinant of human capital levels, we would also expect to find a correlation between wealth and honesty, again with the sign depending on the sign of $\theta'(k)$.

4. Conclusions

It has been argued that an important dimension of the problem of opportunism is the possibility that some agents will act opportunistically less often than others because such behavior has been inculcated in them from childhood. We have shown why human beings may have evolved the capacity for such moral behavior and why we may be flexible enough to change our behavior from one generation to the next in response to incentives.

It follows that it is meaningful to speak of honesty as a form of social capital that can be accumulated. Since honesty affects output, which affects physical capital accumulation, and

[12] This result also holds for the endogenous growth case. What matters for $\theta(k)$ is the firm's level of capital-per-worker since it is the firm's capital stock that will largely determine how much damage a shirking worker will do. Thus, the external effect of aggregate capital factors out above and does not affect the result.

since capital intensity affects the returns to honesty, capital and honesty are co-determined in the long run. We have examined the factors that affect the accumulation of honesty and capital in both exogenous and endogenous growth models. We show that the exogenous growth model may exhibit multiple stable steady-state levels of per-worker-output and honesty if θ, the proportion of output lost due to shirking by a dishonest worker (a D) that was appropriated by the D, decreases with capital accumulation. The endogenous growth model leads to a unique steady-state growth rate of per-worker-output with an associated unique steady-state level of honesty.

The inefficiency and corruptibility of a government can adversely influence the accumulation of honesty in civil society. It can lead to a fall in honesty for each level of the capital stock. Such variation in incentives to be honest can, in an exogenous growth model, lead to the appearance or disappearance of multiple equilibria in output and honesty. Countries with weak and corruptible governments could show negative growth rates and declining levels of honesty as they converge to a low steady state. Therefore, small shifts in the extent of honesty due to variations in government corruptibility can lead to large effects on per-capita-output in the long run. This is also true if there is endogenous growth, for a different reason: the level of honesty then affects the long-run growth *rate* of per-capita-output. However, in the endogenous growth case, small shifts in honesty will not lead to large changes in the extent of honesty in the long run.

If, as seems plausible, θ decreases with capital accumulation, then higher levels of capital are likely to be associated with higher levels of honesty. If there is endogenous growth, this assumption implies growth rates will increase towards a steady state, while the reverse assumption (10) implies that growth rates will decline towards a steady state.

The relation between honesty and human capital was examined. It was shown that if (9) holds, then those with more human capital will be more likely to be loyal to their employers than those with less, while if (10) holds, then the reverse will be true.

We have referred to some empirical cross-country studies of trust (La Porta et al., Fisman and Khanna) that use survey data. The variability in trust that arises in these studies may be due to a combination of factors. In particular, it may be due to differences in incentives, as in repeated-game (Kreps, 1990) and principal–agent theories, as well as due to differences in the proportions of honest types of people, as in the theory advanced here. To distinguish between these different effects and assess their importance, and to test the predictions of our theory more accurately, it would be desirable to measure honest behavior directly via experiments in which the incentives people face are controlled. Much honesty has been found in such experiments, for example, in public good and prisoner's dilemma games in which non-binding promises are made (Ledyard, Frank et al., Brosig). Agency and repeated-game theories cannot explain this behavior. It would be interesting to perform comparable experiments across and within countries to measure variations in such 'intrinsic' honesty. This remains to be done.

Further research may explore the relation between levels of honesty and the structure of contracts and the organization of firms. Because the nature of contracts offered by any given firm affects the accumulation of honesty in the entire economy, there is no reason to suppose that privately optimal contracts will be socially optimal. Consider the example given in Section 1. Notice that the example discussed only those honest types (H's) that were indistinguishable from dishonest types (D's). Suppose h is only slightly higher than

the threshold at which a firm would switch from offering a single wage, to conditioning the wage on revenue. With perfect insurance D's get higher payoffs due to shirking. With the other contract, both types get equal payoffs, thus discretely increasing the incentive to accumulate honesty (which is socially desirable) at very low cost to the firm. (When h is near the threshold, then from the firm's point of view, the full insurance contract is only slightly better than the other one.)

The relation between monitoring, labor-market structure, and efficiency also remains to be investigated. Suppose firms can incur expenditure (real resources) on monitoring workers whose honesty is unknown, and thus reduce losses due to shirking and corruption of various kinds. If the labor market is not very competitive so that firms are local monopsonies, firms may monitor too little because they do not take into account the long-run effect on the proportion of honest workers that more monitoring will have. This can happen because the next generation of workers may not be as locked in to working for the same firm as is the current generation. On the other hand, if the labor market is competitive in the sense that workers are perfectly mobile, then firms may monitor too much with the intention of driving their dishonest workers out to work for other firms. With monopsonistic competition, whether there is too much or too little monitoring will depend on which effect dominates.

Other issues that are of interest include more on the details of the socialization process and how firms can try to socialize their employees to their advantage. The possibility that firms may be opportunistic has been ignored in this paper since it is assumed that this will be detrimental to their long-run interest. However, this may not be the case in the presence of shocks or if the discount rate is not low enough. Thus, how firms will be organized to provide assurance to employees and when firms will be controlled by honest or dishonest persons remain open questions.

Acknowledgements

We thank Abhijit Banerjee, Samuel Bowles, Michael Jerison, Joel Schrag, Rajiv Sethi, William Shughart, Rohini Somanathan, and two anonymous referees for helpful comments.

References

Akerlof, G.A., 1983. Loyalty filters. American Economic Review 73, 54–63.

Barkow, J.H., Cosmides, L., Tooby, J. (Eds.), 1992. The Adapted Mind. Oxford University Press, New York.

Bowles, S., 1998a. Mandeville's mistake: the evolution of norms in market environments. Mimeo, University of Massachusetts at Amherst.

Bowles, S., 1998b. Endogenous preferences: the cultural consequences of markets and other institutions. Journal of Economic Literature 36, 75–111.

Brosig, J., 2002. Identifying cooperative behavior: some experimental results in a prisoner's dilemma game. Journal of Economic Behavior and Organization 47, 275–290.

Coles, R., 1997. The Moral Intelligence of Children. Oxford University Press, Oxford.

Cosmides, L., Tooby, J., 1992. Cognitive adaptations for social exchange. In: Barkow, J.H., Cosmides, L., Tooby, J. (Eds.), The Adapted Mind. Oxford University Press, New York, pp. 163–228.

Erard, B., Feinstein, J.S., 1994. Honesty and evasion in the tax compliance game. Rand Journal of Economics 25, 1–19.

Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: experimental evidence. Econometrica 65, 833–860.

Fehr, E., Gächter, S., 2000. Fairness and retaliation: the economics of reciprocity. Journal of Economic Perspectives 14, 159–181.

Fisman, R., Khanna, T., 1999. Is trust a historical residue? Information flows and trust levels. Journal of Economic Behavior and Organization 38, 79–92.

Frank, R.H., 1987. If *Homo economicus* could choose his own utility function, would he want one with a conscience? American Economic Review 77, 593–604.

Frank, R.H., 1988. Passions Within Reason: The Strategic Control of the Emotions. W.W. Norton, New York.

Frank, R.H., Gilovich, T., Regan, D.T., 1993. The evolution of one-shot cooperation: an experiment. Ethology and Sociobiology 14, 247–256.

Grossman, H.I., Kim, M., 2000. Predators, moral decay, and moral revivals. European Journal of Political Economy 16, 173–187.

Güth, W., Kliemt, H., 1994. Competition or cooperation: on the evolutionary economics of trust, exploitation, and moral attitudes. Metroeconomica 45, 155–187.

Hirshleifer, J., 1987. On the emotions as guarantors of threats and promises. In: Dupré, J. (Ed.), The Latest on the Best: Essays in Evolution and Optimality. MIT Press, Cambridge, pp. 307–326.

Kagan, J., Lamb, S. (Eds.), 1984. The Emergence of Morality in Young Children. University of Chicago Press, Chicago.

Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. Quarterly Journal of Economics 112, 1251–1288.

Kreps, D.M., 1990. A theory of corporate culture. In: Alt, J., Shepsle, K.J. (Eds.), Perspectives on Positive Political Economy. Cambridge University Press, Cambridge, pp. 90–143.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R.W., 1997. Trust in large organizations. American Economic Review 87, 333–338.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–194.

Mayr, E., 1997. This is Biology. Harvard University Press, Cambridge.

Noe, T.H., Rebello, M.J., 1994. The dynamics of business ethics and economic activity. American Economic Review 84, 531–547.

Romer, P.M., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.

Rubin, P.H., Somanathan, E., 1998. Humans as factors of production: an evolutionary analysis. Managerial and Decision Economics 19, 441–455.

Sen, A.K., 1979. Personal utilities and public judgements: or, what's wrong with welfare economics? Economic Journal 89, 537–558.

Trivers, R.L., 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46, 35–57.

Wilson, J.Q., 1993. The Moral Sense. Free Press, New York.