Market Incentives and the Evolution of Intrinsic Motivation*

Timothy Besley
LSE and CIFAR

Maitreesh Ghatak
LSE

February 2016

Abstract

This paper introduces a model where the extent of intrinsic motivation co-evolves with the structure of rewards in society through workplace socialization. Workers vary in terms of intrinsic motivation, and firms compete for workers. Some firms rely on intrinsic motivation and others on extrinsic rewards in equilibrium. Cultural change arises through workers being influenced by social interactions with co-workers depending on their types and the reward structure. A society can become more or less intrinsically motivated depending on endogenously-determined reward structures and initial conditions. We show that in the long-run economies that rely on intrinsic motivation or extrinsic incentives will have similar levels of average productivity but the former will have higher welfare levels.

*We are grateful to workshop participants at the LSE and the conference on Dysfunctional Institutions at the World Bank for helpful feedback on an earlier version of this paper.

1 Introduction

Economists are now paying increasing attention to culture and how it affects economic outcomes, whether it is through peer effects, social norms, values, or institutions.[1] Such investigations emphasise that culture is not exogenous. This follows studies in anthropology and sociology which emphasises that cultures evolve along with their surrounding environments such as the perception of threat or economic circumstances.[2] More specifically, in the context of economic development, standard economic approaches have homed in on factors such as capital accumulation and technology as the prime drivers of change taking cultural features as given. This contrasts with some of the most influential sociological thinkers such as Durkheim (1893) and Polanyi (1944) for whom transformation in culture is a *sine qua non* of economic change. Development of markets can, for example, change the payoffs of agents with different types of preferences, and through social interactions, this can lead to adaptive preference change as individuals' types evolve in response to the enviornment in which they operate. Over time, this can affect the operation of the economy, for better or worse.

In this paper, we are interested in the two-way interaction of this process: how culture shapes economic outcomes and how economic outcomes in turn affect culture. It explores this idea in a specific context by looking at the evolution of intrinsic motivation in response to the incentives which firms use in a market economy. We model this as preference change which affects effort incentives and hence productivity. A society with more intrinsically motivated

[1] See Bisin and Verdier (2011) and Fernandez (2011) for reviews. There is now a large literature on peer effects at the micro-level in schools, neighbourhoods, and workplaces, on individual behaviour relating to education, consumption behaviour, as well as productivity. See, for example, Bramoullé et al (2008), and Mas and Moretti (2009). There is also an emerging literature on how culture matters at a more macro-level, via institutions and social norms, such as Alesina and Sapienza (2015) or through values that are conducive to fostering innovation and growth (e.g., Gorodnichenko and Roland, 2011).

[2] See Gelfand et al (2011) for a thirty-three nation experimental study. Of course, this has been a core issue in social theory from the earliest sociological enquiries of Auguste Compte and running through the ideas of other key early figures in the development of sociological ideas such as Herbert Spencer, Emile Durkheim, and Max Weber.

workers can exploit this by relying less on material incentives to motivate its workers. In our framework, the equilibrium incentives that emerge affect the "fitness" of each preference type creating a process of change where preferences and incentives coevolve. We will study the limit points of this process and consider whether the extent of intrinsic motivation increases or decreases over time.

A key feature of our approach is to study how reward structures affect cultural dynamics. The approach taken here emphasizes the importance to this process of heterogeneous reward structures across firms which we see in reality; some occupations, such as the financial sector, make heavy use of bonus pay whereas others such as academia and the medical profession do not.[3] There is a variety of explanations for this which include the extent of measurability in outputs as emphasized, for example, by Holmström and Milgrom (1991). However, it could also be explained by the extent of intrinsic motivation varying by sector/task. For example, it could depend upon how pro-social an output is and how each worker values this pro-social dimension. In Akerlof and Kranton (2000, 2010), this process is rooted in fluid identities which shape motivation. Besley and Ghatak (2005) focus on agents who are heterogeneous in terms of their intrinsic motivation they derive from their performance, depending on how aligned their preferences are with the mission of organizations they work for. Here, we emphasize how reward structures in firms will change over time as the proportion of motivated workers changes in response to workplace socialization.

We mainly interpret our approach through the lens of the intrinsic motivation literature in psychology such as Deci and Ryan (1985). They emphasise how intrinsically motivated workers who are subject to extrinsic rewards and monitoring can become demotivated. However, more in line with the historical sociological literature, our approach could also be interpreted in terms of reciprocal behavior in work environments. An employer grants a worker autonomy to perform a task without being monitored and a worker repays this trust by putting in effort. This corresponds to a particular view of traditional employ-

[3] See Gittleman and Pierce (2013) for a discussion of some core facts for the U.S..

ment relations which dominated pre-capitalist societies where work was often allied to social relations rather than payment for an end product. As capitalist development proceeded then so did the use of reward structures based on monitoring and output-based rewards with less focus on reciprocity and trust. This view of social and economic change has been emphasized by sociologists such as Durkheim (1893) and Polanyi (1944) who see changes in the nature of the employment relationship as one the central cultural processes which evolved with the advent of a market economy. Along the way, Durkheim emphasized how workers who expected to work on the basis of trust would experience psychic costs which he referred to as *anomie*. In fact there is a wider psychological appreciation of this following the literature on cognitive dissonance following Festinger (1957). Workers who believe that employment relations should be trust-based suffer a loss in well-being from being employed in an environment where they are monitored and/or rewarded based on their output.

In our model of change, the extent of intrinsic motivation coevolves with the structure of rewards in society in response to work-place socialization. Specifically, the extent of selfishness or intrinsic motivation in the population evolves jointly with rewards. Cultural change arises in response to workers being influenced by the rewards that different types receive as well as social interactions with co-workers. A society can become more selfish or more intrinsically motivated depending on endogenously-determined reward structures determined when firms compete for workers. Some firms will rely on intrinsic motivation and others on extrinsic rewards which will evolve over time. We emphasize the role of the underlying technology in shaping this path. Moreover, there is hysteresis in the system which determines the path that an economy follows.

From a theoretical point of view, a key result is that a competitive economy with heterogeneous motivation and team production cannot support a separating equilibrium where intrinsically motivated agents sort into homogenous workplaces. This is important in our context since it is this mixing of different types which fosters cultural change. Fully segmented economies would not experience the dynamics that we study without invoking random mutations

in populations.

Finally, we show that why societies with more intrinsic motivation may foster greater worker well-being without necessarily having a detrimental effect on productivity. Yet, this does not guarantee that intrinsic motivation persists over time. Indeed, it provides an underpinning of recent critiques of market economies, such as Sandel (2012), who raises the concern that market-based reward structures can change the values of market participants in detrimental ways. We will also show how productivity shocks and assortative migration can influence this possibility.

The remainder of the paper is organized as follows. In the next section, we discuss some related literature. Section three sets up the core model and section four looks at its dynamic properties, both positive and normative. Section five draws out some implications and section six contains concluding remarks.

2 Related Literature

Below we discuss two literatures that are most relevant to this paper, on intrinsic motivation and cultural evolution, in greater detail.

Intrinsic Motivation Ryan and Deci (2000) suggest that motivation comes in four different varieties that can be mapped into the approach taken here. At one extreme (external regulation) is pure externally motivated rewards as in the standard economic model discussed above. Next to that is behavior that is motivated either by self-image or impressing others (introjection). In neither of these cases is an activity valued for its own sake. In models of identification, an agent comes to value an action and endorses the goals associated with the task. Finally they propose integration where the agent's preferences are congruent with the task in hand. Then intrinsic motivation is a residual category, with inherent enjoyment and satisfaction from the task or its outcome driving an agent to act.

A range of psychological experiments have looked at these different forms

of motivation. Drawing on evidence from these, Ryan and Deci (2000) emphasize two main factors which encourage intrinsic motivation: (i) competence, i.e. individuals being more motivated by performing tasks they are good at and (ii) autonomy, i.e. having freedom of choice over aspects of how the task is performed. The latter is connected to the idea that workers who are trusted to behave without being monitored are more likely to act with intrinsic motivation.

Economists have studied the origins and implications of intrinsic motivation in economic settings.[4] A first-order effect of intrinsic motivation is to reduce the need for explicit incentive pay. If individuals are given the autonomy to perform tasks where they have competence, then they will be more productive according to the experimental findings. The idea is that individuals will donate their effort "for free" rather than because they are paid to do so. So, in principle, such donations diminish the classic effort-based agency problem, enhancing the scope of the division of labour and increasing productivity in organizations. Starting with this observation, Besley and Ghatak (2005) show that there is a selection argument that follows from it - to the extent that intrinsic motivation varies by worker, organizations that employ more intrinsically motivated work force are likely to have higher levels of productivity. In other words, it is possible that due to this selection effect, use of incentive pay and productivity could be negatively correlated. Viewed from this angle, the use of incentive pay would be considered a symptom of a situation where the agent does not have enough intrinsic motivation.

In their interpretation of intrinsic motivation, Benabou and Tirole (2006, 2011) argue that self-image is also important as a motivator; individuals need not only prove things to others but also to themselves. Individuals may have a sense of the kind of person that they want to be and may want to prove to themselves, via their choices, that this is who they are. In their model, which actions individuals choose will depend on the signals that they emit are perceived by others. There is evidence from various experiments that individuals do not act in selfish or opportunistic ways even in anonymous,

[4] See Frey (1997) for an early discussion.

one-shot interactions.

This approach has the attractive feature that it can provide a way of exploring the question of whether conventional monetary incentives crowd in or crowd out non-pecuniary motivation. In a well-known experiment (see Deci (1975)), college students were either paid or not paid to solve an interesting puzzle, and it was found that those who were not paid spent more time on it and also reported greater interest in the task. In the framework of Benabou and Tirole (2003), a worker may respond negatively to a task for which he is offered a higher reward since he may infer from this that the task is less likely to be one that he values or he is good at. Another argument that is often used in this context is that rewards discourage creativity and risk-taking (Kohn, 1993).

The reputational approach has similarities with the identity-approach of Akerlof and Kranton (2010) who argue that people are moved to act because they associate a particular way of behaving with adopting a particular identity. Such identities are objects of choice and particular "ideal types" are created to which people may aspire. Individuals get utility both from the act itself and any rewards that it brings and how the act conforms or contradicts the identity that the person aspires to. Moreover, this can change over time and may vary according to location and culture. Akerlof and Kranton (2010) suggest that conventional economic approaches which focus on pay-for-performance are likely to lead to wasted effort in situations where a weak sense of identity with the tasks assigned is the cause of organizational failure. Such ideas have been influential in the organizational sociology literature following on the analysis of bureaucracy in Weber (1922).

Socialization and Cultural Evolution Our primary interest here is in how motivation evolves over time and responds to socialization. The approach that we take builds on models of cultural evolution as developed in anthropology by Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985).[5] The cultural evolution paradigm has been important in exploring the

[5] The literature on cultural evolution is surveyed in Bisin and Verdier (2011).

basis of unselfish behavior in either kin groups of wider social groups. They allow humans to "inherit" their behavior based on formal rules. In the classic example studied in this literature, there is public goods game with a free-rider problem where rational self-interested individuals will take advantage of others. There are three mechanisms of change: mutations, genetic drift (in finite populations) and natural selection. Boyd and Richerson (1985) studies models of "conformist transmission" where individuals are assumed to be disproportionately likely to imitate the more common behavioral types among their cultural parents. This creates a force that increases the frequency of more common types in the population.

Those who have studied cultural evolution vary in the importance that they attach to the different forms of social interactions as the conduit for social influence. Here, our focus is on socialization in the work place as an example of socialization through the life cycle. However, there are many forms of social interaction which can serve as the basis for evolutionary change.[6] Much of the work among economists such as Fernandez (2011) and Guiso et al (2006) has focused on intergenerational transmission.[7] They find strong empirical evidence of cultural influence across generations.

The model developed here shares the core structure of population dynamics with these papers. However, in common with economic approaches, it puts payoffs at the heart of the process which are endogenously determined by behavior. This corresponds to the indirect evolutionary approach introduced in Güth and Yaari (1992) and Güth (1995). This approach is discussed in the context of collective action in Ostrom (2000). Most of that literature has focused on small-scale cooperation.

There is a small literature in economics which has looked at socialization of preferences. Unlike the models in anthropology, these have tended to model this as strategic behavior of parents towards their children. For example, Bisin and Verdier (2001) develop a model where the decision to socialize children

[6] See also Akerlof and Kranton (2010) for discussion of the evolution of identity as a form of cultural transmission.

[7] This is also the approach taken in Besley (2015) and Bisin and Verdier (2001).

is strategic and depends on the payoffs that the children will receive weighed against the "social distance" that it creates between parents and children. Tabellini (2008) uses a related approach to look at the evolution of preferences and cooperation.[8]

The paper is also related to a body of classical sociological literature on socialization and cultural change. These are most associated with social scientists such as Durkheim (1893), Merton (1968) and Polanyi (1944) for whom the emergence of a market economy also leads to changes in social structure, and cultural norms that co-evolve with economic change.

The idea of socialization is fundamental in sociology which has developed elaborate theories of this process. An important distinction is between primary socialization which takes place in families and as part of the parenting process with secondary socialization which occurs in other forms of social groups and can evolve over the life cycle, even into old age. One key example of the latter and the focus of this paper is on workplace socialization. According to Van Maanen and Schein (1979):

> "organizational socialization refers ... to the fashion-in which an individual is taught and learns what behaviors and perspectives are customary and desirable within the work setting as well as what ones are not." (page 4)

From the start, organizational psychologists have emphasized the importance of group dynamics in shaping cultural change (see Schein, 1965). The key observation that we make use of in our paper is that motivation is not fixed but is fluid and responsive to the environment to which individuals are exposed and can be a source of social and economic change.

3 The Model

The economy comprises a range of firms and workers who produce in teams of two. Each period, firms and workers match. The key feature of the

[8] See Bisin and Verdier (2011) for an overview of this literature.

model which drives the dynamics is that there is no equilibrium with perfect matching where intrinsically motivated workers work together in the same firms. The static equilibrium of the labour market involves workers with heterogeneous motivation to match together, which is due to the asymmetric information about worker motivation. It is this mixing of workers of different types that drives the socialization process and the evolution of the distribution of motivation over time.

Framework The production technology requires two workers to produce output in any firm.[9] Firms use team production and a unit of output is produced only if *both* agents put in effort. When this happens, the firm owner gets an output of value π where heterogeneity π is either technological or based on differences in human capital among firm owners. If only one worker puts in effort, then output is normalized to be 0.

Workers can choose to put in one unit of effort $e \in \{0, 1\}$. The effort choices of workers are unobservable to firms but the output is verifiable and hence bonus contracts can depend upon it. There are two types of workers denoted $\tau \in \{m, s\}$ where m stands for motivated and s for selfish. Effort costs c to a selfish agent. Intrinsically motivated agents not only incur no cost from putting in effort, they enjoy it and receive a non-pecuniary payoff of $\theta > 0$. A firm pays a flat wage to workers, and may also pay a bonus contingent on output. Selfish workers, assuming their incentive constraint is satisfied, may put in effort when there is incentive pay. However, if motivated workers are paid for their effort, they incur a non-pecuniary cost of $\gamma > 0$ which is subtracted from the payoff θ, as they prefer a trust-based reward system as opposed to one that is output-based. Motivated workers *always* put in effort whether they are paid or not because choosing zero effort conditional on being employed gives them a disutility $\delta < 0$, where $\theta - \gamma \geq \delta$.

There is a measure $\frac{N}{2}$ of producers (firms) and a measure 1 of workers. Given the production technology, the number of jobs is N. We restrict at-

[9] We could easily add an "intercept" level of output which is generated without any effort. This would not materially change the results in the paper.

tention to the case where workers are on the long-side of the market, i.e., $N < 1$.

Workers who are not employed have a subsistence consumption level of $z \geq 0$, which also becomes the outside option of an employed worker. Time is indexed by t, and we look at an infinite time horizon as far as the economy is concerned. Let μ_t be the fraction of motivated workers in the population at date t which is assumed to be common knowledge across firms and workers.

We restrict attention to firms whose productivity π is continuously distributed in the interval $[2(c+z), \Pi]$ where $\Pi > 2(c+z)$. The assumption $\pi \geq 2(c+z)$ implies that all firms will choose to be active in the market equilibrium that we describe below. The distribution of the type of a given firm, π, is given by the CDF $G(\pi)$, and corresponding PDF $g(\pi)$, both assumed to have the standard properties.

The relationship between workers and firms lasts one period. We can either think of workers and firms both being active for one-period and then being replaced by a distribution that is identical in all respects except for the fraction of motivated workers μ_t. In this case, workplace socialization affects the intrinsic motivation of the next cohort of workers through an additional link of inter-generational transmission of values within the family. Alternatively, we can think of workers being long-lived but firms being short-lived, so that workplace socialization directly affects the intrinsic motivation of the same cohort of workers the next period. Either way, we are assuming away any market dynamics due to evolving worker reputations. We will discuss how this might affect our results below.

Workers are randomly matched with firms who post employment contracts which comprise a type-specific wage, w_τ, and an output contingent payment (bonus) b_τ The type of a worker is private information as far as firms are concerned. Effort is not contractible and workers have no wealth which they can post as a bond against poor performance. We assume that there is a limited liability constraint which takes the form of a minimum consumption constraint: workers must be paid up front a flat payment of at least z. Also, bonuses have to be non-negative. Given that workers are on the long-side of

the labour market, we will suppose the outside option is unemployment where a worker receives a subsistence consumption level of z.

We consider equilibrium contracts $\{w_m, b_m, w_s, b_s\}$ in a market equilibrium where firms compete for workers. While in principle, we should allow these contracts to be indexed by t, we will show that the same contracts will be offered at every date and hence ignore this complication up front.

The model has both adverse selection and moral hazard. We will require that contracts are incentive compatible in two senses (i) workers select the contract intended for their type and (ii) effort decisions are optimal for selfish types (recall that motivated workers put in effort unconditionally).

Benchmark: The Full Information Allocation As a benchmark let us describe the efficient allocation under full-information. Given that $\pi \geq 2(z+c)$ and production requires both workers to put in effort, $e_m = e_s = 1$ is the efficient allocation. A motivated worker should be given no bonus, so that he gets a non-pecuniary utility of θ from working and a flat wage of at least z. The selfish worker would have to be given a compensation of at least $z+c$. If there are enough motivated workers to fill all jobs, then only they will be hired in equilibrium as they have the same productivity but a lower cost of effort than selfish workers. However, if $\mu < N$ then some selfish workers would be hired too. However, they would be less profitable for firms.

Imperfect Information We now consider what happens in a world of imperfect information (moral hazard and adverse selection) with competitive matching of workers to firms.

Effort Decisions Let $E(b, \tau)$ be the effort decision of type τ when the bonus is b. By assumption, $E(b, m) = 1$. Consider the behavior of selfish agents. We make the standard tie-breaking assumption that if a selfish agent is indifferent between supplying effort or not, he supplies effort. We have the following useful preliminary result, whose proof, along with the other formal proofs, are in the Appendix:

Lemma 1: *If $b \geq c$, then there is an equilibrium in which all selfish agents put in effort whether they are matched with a selfish or a motivated agent. For $b < c$ selfish agents do not put in effort.*

We will henceforth focus on the equilibrium where:

$$E(b,s) = \begin{cases} 1 & \text{if } b \geq c \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

This is the risk-dominant as well as payoff-dominant equilibrium in our context.

Using this, we can define the payoffs of each type. The expected effort-related payoff component of a selfish agent is:

$$V(b,s) = E(b,s)[b-c],$$

while for a motivated agent it is:

$$V(b,m) = \begin{cases} \theta & \text{if } b = 0 \\ \theta + b - \gamma & \text{otherwise.} \end{cases}$$

Self-Selection We will consider incentive compatible allocations where workers self-select. The conditions for this are standard:

$$\begin{aligned} w_s + V(b_s,s) &\geq w_m + V(b_m,s) \\ &\text{and} \\ w_m + V(b_m,m) &\geq w_s + V(b_s,m). \end{aligned} \tag{2}$$

These simply say that each type should pick the contract intended for them.

Minimum Consumption Constraint of Workers We assume workers have a minimum consumption constraint that requires an up-front wage

payment of at least z and the limited liability constraint implies that bonuses have to be non-negative:

$$w_\tau \geq z \text{ and } b_\tau \geq 0. \tag{3}$$

These constraints restrict the ability of firms to extract the utility rent θ that motivated workers earn.

The Outside Option of Workers The outside option of a worker is unemployment, where he receives a subsistence consumption level of $z > 0$. This implies that total remuneration cannot fall below z, i.e., $b_\tau + w_\tau \geq z$. The limited liability and minimum consumption constraints (3) together ensure that this constraint is satisfied We assume that there is a very small disutility $\varepsilon > 0$ from being unemployed so that all workers strictly prefer to work if they can even if the consumption level if z under both options. Hence the outside option for both types of worker is $z - \varepsilon \geq 0$. Let the outside utility of workers be denoted by $\{u_m, u_s\}$. Voluntary participation requires that

$$\bar{u}_\tau \geq z - \varepsilon. \tag{4}$$

Competition and Profit Maximization Contracts must also be consistent with competitive profit maximization by firms. Firms offer a common contract $C = \{w_m, b_m, w_s, b_s\}$ and are atomistic and take the outside utility of workers, denoted by $\{u_m, u_s\}$, as fixed. We will require that no firm can deviate from the proposed equilibrium contract and attract workers in a way that satisfies the self-selection, effort incentive-compatibility for selfish workers, the limited liability and minimum consumption constraints (as well as the participation constraints although these are automatically satisfied whenever the last set of constraints are).

Formally, suppose that a firm hires two workers i and j drawn from the set of all workers, denoted by $\mathcal{S}$. Let they be of type $\tau(i)$ and $\tau(j)$. Let $\mathcal{P} \equiv \{(s,s),(s,m),(m,s),(m,m)\}$ be the set of permutations of the types of worker pairs with typical element p. When we consider $\{\tau(i), \tau(j)\} \in \mathcal{P}$ we cover all permutations of worker types that a firm could choose from.

Focusing on contracts where (1) and (2) hold then the profits of the firm are:

$$\Pi\left(i,j:\pi,C\right)=E\left(b_{\tau(i)},\tau\left(i\right)\right)E\left(b_{\tau(j)},\tau\left(j\right)\right)\left[\pi-b_{\tau(i)}-b_{\tau(j)}\right]-w_{\tau(i)}-w_{\tau(j)}.$$

Given any equilibrium contract C, the equilibrium utilities of workers are $w_s+V\left(b_s,s\right)$ and $w_m+V\left(b_m,m\right)$.

For profit maximization we require that, for all $\{i,j\}\in\mathcal{S}$ there does not exist $C'=\left\{w_m^{'},b_m^{'},w_s^{'},b_s^{'}\right\}$ which satisfies (1), (2) and (3) such that:

$$w_s^{'}+V\left(b_s^{'},s\right)\geq w_s+V\left(b_s,s\right)\text{ and }w_m^{'}+V\left(b_m^{'},m\right)\geq w_m+V\left(b_m,m\right)\quad(5)$$

and $\Pi\left(i,j:\pi,C'\right)>\Pi\left(i,j:\pi,C\right)$. This says that there is no profitable deviation available to any firm that can attract workers satisfying the self-selection, incentive-compatibility and limited-liability constraints, and would strictly increase its profits.

Timing The timing of the model is as follows.

1. Firms post contracts $\{w_m,b_m,w_s,b_s\}\in C$.
2. Firms and workers match, workers choose their effort levels, and output is realized.
3. Socialization in the work place leads to the fraction of motivated workers in the economy being updated from μ_t to μ_{t+1}.

We will work backwards through each stage of the model.

Socialization and Cultural Dynamics If a worker is unmatched his type does not change. Given a set of equilibrium contracts C^* and a fraction μ of motivated workers, for workers who find employment let $U\left(C^*,\mu,\tau\right)$ be the expected utility of being a type $\tau\in\{m,s\}$ and let

$$\Delta\left(\mu\right)=U\left(C^*,\mu,m\right)-U\left(C^*,\mu,s\right)$$

be the utility difference between the motivated type and the selfish type. We will characterize $\Delta(\mu)$ below.

Socialization will take place by exposure to co-workers who serve as "cultural parents" in this setting. We assume that even if a worker type is not known to the firm, in the process of working together as a team, co-workers find out the type of each other. They are also assumed to observe the contracts received by each other, as well as the output, and so they can infer the payoffs they receive (both *ex ante* as well as *ex post*). We assume workers are influenced by *ex ante* payoffs as that is what would be relevant in the next period (for them, or their offspring).

How the process of socialization works depends on the nature of workplace mixing. We assume workers do not interact with each other before and therefore, the type of a co-worker is randomly drawn given the contracts on offer and μ_t, given the nature of the matching process between firms and workers.

We will suppose that socialization has bite only in situations where co-workers are heterogeneous and is based on the "fitness" of each type as defined by $\Delta(\mu)$. In a world where a particular type earns higher returns, it has greater fitness and there will be a move towards more of that type emerging in the population. Specifically, let

$$\rho(\Delta(\mu_t)) = \frac{\exp[\Delta(\mu_t)]}{1+\exp[\Delta(\mu_t)]}$$

be the probability a selfish type becomes a motivated type if he is mixes with a motivated type in a firm.[10] The complementary probability of becoming selfish from motivated is $1-\rho(\Delta(\mu_t))$ and so whether $\rho(\Delta(\mu_t))$ is greater or less than one half determines the direction of the dynamics. Observe that $\rho(0) = 1/2$ and $\rho'(\Delta) > 0$. The particular functional form of the probability distribution (logistic) is not important for our analysis. This is a specific way

[10] Note that it would straightforward to allow for cultural bias with

$$\rho(\Delta(\mu)) = \frac{\exp(\beta+\Delta(\mu))}{1+\exp(\beta+\Delta(\mu))}$$

where $\beta > 0$ is a bias towards intrinsic motivation and $\beta < 0$ a bias against.

of capturing payoff monotonic selection dynamics as in Sethi and Somanathan (2001). Here, the utility difference between these types determines the pattern of cultural change in mixed contexts. This is motivated by supposing that any randomly selected agent is motivated if

$$\Delta(\mu_t) + \zeta \geq 0.$$

where ζ has the same distribution as in a standard discrete-choice logit model.

We can write down an equation for the population dynamics of motivated workers as:

$$\mu_{t+1} = \sigma\left\{\mu_t^2 + 2\mu_t(1-\mu_t)\rho(\Delta(\mu_t))\right\} + (1-\sigma)\mu_t \tag{6}$$

where σ is the probability that a worker is randomly matched.

Thus if $\sigma = 0$, there is no socialization at the workplace. Thus, $\sigma > 0$ is needed for cultural evolution.[11] Writing this as a difference equation, we have that:

$$\mu_{t+1} - \mu_t = \sigma\mu_t(1-\mu_t)\left[2\rho(\Delta(\mu_t)) - 1\right].$$

Thus the sign of the change is determined by $\rho(\Delta(\mu_t)) \gtrless 1/2$ or equivalent $\Delta(\mu_t) \gtrless 0$. The extent of socialization in the workplace in the population, σ, affects the "speed" of the cultural dynamics but not its direction. Given that firms and workers match randomly, and a fraction $(1-N)$ of workers remain unmatched, we will see that $\sigma = N$ in equilibrium.

Matching and Effort Choice At stage 2, firms and workers match, workers choose effort levels and, at stage 1, firms choose contracts. We select a specific extensive form for assigning firms to workers. However, all that is required is that all firms have their labor demand satisfied by workers who are assigned in an incentive-compatible way. Given the contracts posted by firms, workers apply to a set of firms posting a specific contract (or a pair of contracts). Given that applying is not costly, while unemployment is costly and all contracts are

[11] This would not be the case if we were to allow for random mutations.

incentive-compatible, we assume every worker of each type applies to every firm, and is chosen at random from among the applicants. We assume that the firms are indexed in a random sequential order and the market for workers is assumed to clear by supposing that firms choose their workers sequentially with firm k moving before firm k' where $k < k'$, and with each worker then retiring from the pool of applicants once it has been matched with a firm. At the end of this process, some workers remain unmatched and engage in the subsistence activity.

If $b_s \geq c$ in a separating equilibrium or $b \geq c$ in a pooling equilibrium, then by Lemma 1, selfish workers supply effort. Motivated workers always supply effort.

Equilibrium Contracts In stage 1, we restrict attention to contracts that satisfy (1), (2), and (3).[12] Failing to satisfying these constraints is inconsistent with profit-maximization (including the ability to recruit workers). We now proceed to characterize equilibrium contracts C^* which satisfy these constraints and are profit maximizing for all $\pi \in [2(c+z), \Pi]$. We will show that C^* comprises two sets of pooling contracts both of which are typically on offer in a market equilibrium.

We will focus on the parameter range where

$$0 > \theta + c - \gamma > -\varepsilon. \tag{7}$$

To interpret this, note that the intrinsically motivated worker receives $\theta - \gamma + c$ if he works and is given an incentive contract that would induce effort from a selfish worker. He would receive $-\varepsilon$ if he is unemployed. So motivated workers would prefer work to unemployment even if they are incentivized. However, the disutility that they get from work is undermined by being incentivized. If instead $\theta + c - \gamma > 0$ then motivated workers always have a higher expected payoff than selfish workers, independent of μ and the terms of their employment. In this case, the distribution of worker motivation will always tend to a

[12] Recall that satisfying the last pair of constraints also means that (4) is satisfied.

fully motivated workforce, which is less interesting than the case we focus on.

Condition (7) will capture what Durkheim (1893) called *anomie.* This is the case where there is no alignment between values and what happens in practice which causes a state of anxiety to workers leading to demoralization. This will hold in our framework as long as γ is large enough. Condition (7) plays a key role is that it means that a motivated worker assigned to a job where she is being incentivized is strictly unhappy relative to a selfish worker doing the same job. Thus, there is a downside to being a motivated worker if one is not matched to a job where that motivation is harnessed rather than being "undermined" by being placed under an incentive contract.

First consider the case where $\mu_t > N$, that is, there are enough motivated workers to fill all jobs. The first-best allocation is clearly not incentive compatible, since if $w = z$ and $b = 0$ is offered, selfish workers would prefer to take this contract than engage in subsistence activity. The only way selfish workers could be discouraged is if the minimum consumption constraint was not binding and w was set at $z - \theta$. By assumption ε is arbitrarily small and so this contract would discourage the selfish workers and motivated workers would still strictly prefer this to the subsistence activity. However, the minimum consumption constraint implies w cannot be set below z and given the surplus of workers, profit maximization implies $w = z$ Now, if in the second-best the contract $w = z$ and $b = 0$ is offered, selfish workers would take this contract as will motivated workers. However, selfish workers would choose zero effort, and a firm of type π will therefore have an expected profit of $\mu^2\pi - 2z$. The other options would be to offer a pair of separating contracts that will satisfy the self-selection constraints for both types of workers or a pooling contract such that selfish workers will put in effort under it.

In the other case where $\mu_t < N$, firms need to hire both motivated and selfish workers for productive activity to be possible. In this instance, the options available for firms is to offer a pair of separating contracts that will satisfy the self-selection constraints for both types of workers or a pooling contract such that selfish workers will put in effort under it. Our main result on the contracting equilibrium is given in:

Proposition 1 *All contracts in C^* set subsistence wages, i.e. $w_m = w_s = z$. For bonuses, the market offers two possible contracts: a bonus contract where $b_s = b_m = c$ and a fixed-wage contract where $b_s = b_m = 0$. Firms choose which contract to offer as follows:*

1. *if $\pi \geq \frac{2c}{1-\mu^2}$ then $b_s = b_m = c$*
2. *if $\pi < \frac{2c}{1-\mu^2}$ then $b_s = b_m = 0$*

The striking feature of this result is that there is no separating equilibrium. Given the slack in the labour market both types of workers have identical outside options. Therefore, if both types of workers put in effort, from a firm's point of view they are equivalent. As a result, competition among profit-maximizing firms implies $b_s + w_s = b_m + w_m$. If this did not hold, then one type of worker (the one with the lower total remuneration) would be strictly preferred to the other. This, coupled with the self-selection constraint restricts the set of different contracts that can be offered considerably. On the other hand, given that the number of workers exceeds the number of jobs, profit-maximizing firms will choose the lowest wages and bonuses consistent with limited-liability and minimum consumption constraints (in addition to self-selection).

The Proposition says that there are two possible pooling contracts on offer by firms and different from high and low productivity firms. If the firm is highly productive relative to the cost of effort, all of its workers receive an incentive contract which guarantees an output of π. Low productivity firms will choose not to incentivize workers and hence will only produce π if they happen to get two motivated workers which happens with probability $(\mu_t)^2$. Since both these are pooling equilibria, some firms will have a mixture of motivated and selfish workers and these are the firms in which on-the-job socialization will, with some probability, lead to changes in worker motivation.[13]

[13] The model is consistent with the positive correlation between subjective well-being and productivity at the firm level (conditional on other factors that influence firm-level productivity) found by Bryson et al (2015). The result above is also consistent with the finding that there is heterogeneity in the use of performance pay across firms within an economy

This result is important for the study of cultural evolution as it says that workers of different types will be mixing in the workplace and hence there is the possibility of one group influencing the other. We now study the implications of this.

4 The Evolution of Intrinsic Motivation

In the equilibrium described in Proposition 1, selfish workers receive a utility of z in all kinds of firms. Motivated workers receive a payoff of $\theta + c - \gamma < 0$ if they are working in firms where $b \geq c$ or θ if they are working in firms where $b = 0$. The probability that any type of worker is employed is N.

This implies that the expected utility of a selfish worker at the matching stage is

$$U\left(C^*, \mu, s\right) = -\left(1 - N\right)\varepsilon,$$

i.e. there is a prospective disutility of being employed with probability $(1 - N)$. From Proposition 1, a fraction $G\left(\frac{2c}{(1-\mu^2)}\right)$ of firms offer the contract $b_s = b_m = 0$ while a fraction $\left\{1 - G\left(\frac{2c}{(1-\mu^2)}\right)\right\}$ of firms offer the contract $b_s = b_m = c$. Therefore, the expected utility of a motivated worker is:

$$U\left(C^*, \mu, m\right) = \Delta\left(\mu\right) - \left(1 - N\right)\varepsilon$$

where

$$\Delta\left(\mu\right) = N\left[\theta + \left(1 - G\left(\frac{2c}{(1-\mu^2)}\right)\right)(c - \gamma)\right] \tag{8}$$

is the utility gain or loss from being a motivated worker. This can be negative since $\theta + c - \gamma < 0$.

---

and that this fraction is not static over time. For the US economy, Gittleman and Pierce (2013) show that there is greater performance pay in the IT and financial industries. They also find that the incidence has fallen particularly since the early 2000s. More generally, there has been a long-term shift of performance pay away from low wage jobs towards high wage jobs. A more complete model would however have to allow for the possibility of heterogeneity in contracting technology and importance of non-contractible effort in different industries.

Using (8),

$$\Delta'(\mu) = -(c-\gamma)\, g\left(\frac{2c}{(1-\mu^2)}\right)\frac{2\mu}{(1-\mu^2)^2}N \geq 0.$$

This is an important complementarity in the model where the payoff to a motivated agent is greater when there are more motivated workers around. This is because firms will then offer more fixed wage opportunities. Moreover since (7) holds,

$$\Delta(0) < 0 \text{ and } \Delta(1) > 0.$$

Therefore, the *sign* of $\Delta(\mu)$ also depends on μ. Specifically define $\hat{\mu}$ from

$$\theta = \left(1 - G\left(\frac{2c}{(1-\hat{\mu}^2)}\right)\right)[\gamma - c].$$

Then $\Delta(\mu) \geq 0$ for all $\mu \geq \hat{\mu}$ and if $\mu < \hat{\mu}$, then $\Delta(\mu) < 0$. In economies that start off with μ below $\hat{\mu}$, motivated workers are worse off than selfish workers and through workplace socialization, this tends to push μ further down over time. Likewise, in economies where μ is above $\hat{\mu}$, motivated workers are better off than selfish workers and through workplace socialization, this tends to increase μ over time. We now turn to formally showing that this threshold level of fraction of motivated workers, $\hat{\mu}$, is a tipping point.

Long-run Dynamics It is now straightforward to characterize the long-run dynamics of the economy where the worker types are endogenous but infinitely-lived. The economy will have a critical threshold, a tipping point, which determines whether intrinsic motivation grows or diminishes in the population over time. For populations where there are few motivated workers, there is a tendency for motivation to diminish over time as the likelihood of being in a firm that does not incentivize workers is low. However, for populations where μ is initially high, there is a larger reward to being the motivated type as more firms choose to operate without incentives. This gives us the following result.

Proposition 2 *For* $\mu_t < \hat{\mu}$, $lim_{t-\infty}\mu_t = 0$ *and for* $\mu_t > \hat{\mu}$, $lim_{t-\infty}\mu_t = 1$.

Thus, the model exhibits multiple long-run steady states where the initial condition matters. An economy that begins with few motivated workers, will tend to make those that it has unhappy, and this will reduce the fraction of firms in the economy which offer fixed wage contracts over time. Thus, bonus pay will tend to increase over time and socialization will reinforce this effect, tending to favor selfish workers.

The tendency of intrinsic motivation to diminish over time in an economy that uses incentive pay is explained in the model by the fact that intrinsically motivated workers feel demotivated by working in firms with bonus pay and tend to lose in the process of cultural evolution since their type has lower "fitness" than the selfish type. This is consistent with the idea that anomie drives the values of a capitalist economy towards a workforce which works only for material reward.

The opposite is true of economies which begin with many motivated workers which will tend to create more fixed wage contracts. This will tend to give an advantage to motivated workers in the happiness stakes, i.e. they have greater "fitness", the cultural evolution process increases the share of workers with intrinsic motivation in the worker population.

Welfare and the Distribution of Rewards Next we consider aggregate welfare as well as the distribution of rewards in this economy and how this changes with the fraction of motivated workers. Looking first at material rewards, observe that some workers will earn a premium for working in the form of incentive pay. Thus there is a fraction, $\left[1 - G\left(\frac{2c}{(1-\mu^2)}\right)\right]$ of the work force which earns c. As the society converges to one of its extremes described above, inequality *between* workers will diminish. Socialization will also diminish over time, with the extent of anomie experienced by intrinsically motivated workers given by:

$$\mu\left[1 - G\left(\frac{2c}{(1-\mu^2)}\right)\right][\theta + c - \gamma].$$

This is the aggregate disutility of workers who are assigned to jobs which make them unhappy. Happiness among workers also increases at either extreme as

the population becomes more homogenous. Therefore, socialization seems to have benefits in terms of lower inequality and higher aggregate worker utility.

In terms of firms' profits, there is a benefit to relying on motivated workers rather than incentives. To see this observe that a firm that has motivated workers earns π while one that has to pay $2c$ in the form of bonuses makes $\pi - 2c$. Owners of firms (capitalists) therefore benefit from having more motivated workers.

Let us define $\overline{\mu}$ and $\underline{\mu}$ such that

$$\frac{2c}{(1-\overline{\mu}^2)} = \Pi \text{ and } \frac{2c}{\left(1-\underline{\mu}^2\right)} = 2(c+z).$$

In particular, $\overline{\mu} \equiv \sqrt{1-\frac{2c}{\Pi}}$ and $\underline{\mu} \equiv \sqrt{\frac{z}{c+z}}$ and as $\Pi > 2(c+z)$, $\overline{\mu} > \underline{\mu}$.

To calculate aggregate social surplus, we have to look at the expected profit of firms and the net surplus of workers. A fraction N of workers are employed. A fraction $(1-N)$ of workers remains unemployed and their disutility of ε must also be included in the calculation of total surplus. Selfish workers get zero net surplus and so their payoff is not calculated. All motivated workers get θ but a fraction of motivated workers (those who work in firms with $\pi \geq \frac{2c}{1-\mu^2}$) get a disutility of γ. There is a measure $\frac{N}{2}$ of firms. Firms whose productivity lies between $2(c+z)$ and $\frac{2c}{(1-\mu^2)}$ have a total expected profit of $\mu^2 \int_{2(c+z)}^{\frac{2c}{(1-\mu^2)}} \pi g(\pi) d\pi$ while firms with productivity lies between $\frac{2c}{(1-\mu^2)}$ and Π have a total expected profit of $\int_{\frac{2c}{(1-\mu^2)}}^{\Pi} \pi g(\pi) d\pi$. When $\mu \geq \overline{\mu}$, all firms stop using bonuses, and similarly, when $\mu \leq \underline{\mu}$, all firms use bonuses. We denote by $E(\pi) = \int_{2(c+z)}^{\Pi} \pi g(\pi) d\pi$ the unconditional average productivity of firms.

Aggregate social surplus S when the fraction of workers is μ is therefore

given by:

$$S(\mu) = \begin{cases} N\mu(\theta-\gamma) + \frac{N}{2}E(\pi) - (1-N)\varepsilon & \text{for } \mu \in [0,\underline{\mu}]; \\ N\mu\left[\theta - \gamma\left\{1 - G\left(\frac{2c}{1-\mu^2}\right)\right\}\right] & \\ +\frac{N}{2}\left\{\mu^2 \int_{2(c+z)}^{\frac{2c}{(1-\mu^2)}} \pi g(\pi)\, d\pi + \int_{\frac{2c}{(1-\mu^2)}}^{\Pi} \pi g(\pi)\, d\pi\right\} - (1-N)\varepsilon & \text{for } \mu \in [\underline{\mu},\overline{\mu}]; \\ N\mu\theta + \mu^2\frac{N}{2}E(\pi) - (1-N)\varepsilon & \text{for } \mu \in [\overline{\mu},1]. \end{cases}$$

Observe that:

$$\begin{aligned} S(1) &= N\theta + \frac{N}{2}E(\pi) - (1-N)\varepsilon \\ &\text{and} \\ S(0) &= \frac{N}{2}E(\pi) - (1-N)\varepsilon. \end{aligned}$$

Thus $S(1) - S(0) = N\theta > 0$.

Aggregate income Y is the sum total of profits and income of workers. Firms that use incentive pay will show higher income for workers but that has to be balanced against the deduction from firms' profits. Ignoring the base pay of z to workers, which the unemployed earn as well, Y is given by:

$$Y(\mu) = \begin{cases} \frac{N}{2}E(\pi) & \text{for } \mu \in [0,\underline{\mu}]; \\ \frac{N}{2}\left\{\mu^2 \int_{2(c+z)}^{\frac{2c}{(1-\mu^2)}} \pi g(\pi)\, d\pi + \int_{\frac{2c}{(1-\mu^2)}}^{\Pi} \pi g(\pi)\, d\pi\right\} & \text{for } \mu \in [\underline{\mu},\overline{\mu}]; \\ \mu^2\frac{N}{2}E(\pi) & \text{for } \mu \in [\overline{\mu},1]. \end{cases}$$

It is clear upon inspection that $Y(1) - Y(0) = 0$. We have therefore proved:

Proposition 3 *In the long-run economies based on intrinsic motivation will have higher welfare and similar income levels to those which rely on incentives.*

So a society with intrinsically motivated workers is welfare superior to one in which all workers are on incentive-based pay. This is because motivation is a source of positive utility which is undermined by the use of incentives. Note that this would be picked up in happiness or life-satisfaction data but not in

GDP in terms of which two economies, one which relies only on incentives and the other only on motivation, would be similarly placed.

As we have already seen, there is no reason for a society to converge towards the outcome where all workers are motivated. Indeed, if the initial condition is that motivation is scarce this is unlikely. Moreover, we could argue that the drift of most modern societies has indeed been towards greater use of incentives. In terms of our model, this implies over time μ falls since motivated workers have lower utility than selfish workers. This in turn makes contracts that do not use incentive pay less attractive to the lower-productivity firms and the process continues until μ goes to zero. What is interesting is, when μ is low enough, welfare goes up locally as μ goes down - this follows from the fact that $S'(\mu) = N(\theta - \gamma) < 0$ for $\mu \leq \underline{\mu}$. On the other hand, if μ is high enough, welfare goes up locally as μ goes up - this follows from the fact that $S'(\mu) = N\theta > 0$ for $\mu \geq \overline{\mu}$. Even though welfare is locally increasing in μ in the high motivation steady state, while it is decreasing in μ in the low motivation steady state, from Proposition 3 we know an economy that relies on incentive pay only has *lower* levels of welfare than an economy where people work on the basis of intrinsic motivation. That is, the two steady states can be unambiguously ranked in terms of welfare.

This finding does suggest the possibility that a social planner would find a way of steering an economy towards the long-run outcome where an economy relies on intrinsic motivation. The planner could tax firms that rely on incentives as a means of increasing the fraction of opportunities for motivated types. This could see an economy flipping beyond the tipping point of $\hat{\mu}$ and hence getting on a trajectory where μ converges to one.

5 Further Implications

In this section, we draw out two further implications for the framework. We first consider comparison across economies and how differences in productivity affect the evolution of motivational preferences. Second, we look what happens if there are some mobile citizens who can choose where to work.

Structure of Production The threshold level of motivation is dependent upon the distribution of productivity across firms. Thus, consider two economies A and B where productivity differs. Then we have:

Proposition 4 *Consider two distributions of productivity in economies A and B where the first dominates the second in a first order sense, i.e.*

$$G^A(\pi) \leq G^B(\pi) \text{ for all } \pi \in [2[c+z], \Pi].$$

The critical threshold fraction of motivated individuals for economy A, $\hat{\mu}^A$ will be everywhere above the critical threshold fraction of individuals in economy B, $\hat{\mu}^B$.

The more productive economy is therefore strictly less likely to foster a long-run equilibrium with intrinsic motivation. This result is due to highly productive firms wishing to use bonus contracts which diminish labor market opportunities suited for intrinsically motivated types.

This observation allows us to think about the impact of technological improvements on the extent of intrinsic motivation. Polanyi (1944) put forward a view of modernization following economic growth as transforming the values of human societies. He argued that the industrial era led to a "market society" where individuals came to resemble rational egoists. Development had led to the destruction of the basic social order that had reigned because of pre-modern human nature and that had existed throughout all earlier history. Instead of being allocated according to tradition, or norms of redistribution or reciprocity, labor was now sold on the market at market-determined prices. This, according to Polanyi, played a role in transforming the values and motivation of individuals. Our model can make sense of Polanyi's conjecture if we interpret the motivated types in our model as in a reciprocal relationship with their firm whereby they put in effort in exchange for not being monitored or incentivized, the premise of which is not trusting the worker.

The model can also make sense of Polanyi's ideas by consider an increase in the distribution of productivity that pushes up the threshold $\hat{\mu}$. This can

lead to an economy moving to a downward trajectory in terms of intrinsic motivation. However, there is a key difference between our analysis and what Polanyi claims. It is not the existence of a labor market *per se* that affects motivation but the form that contracts take. Our zero bonus contracts could be thought of as some kind of more reciprocal labor relations. However, their use is also responsive to market conditions. In turn, it affects distribution of preferences between selfish behavior and intrinsic motivation. It is the decline in the latter that can be viewed as consistent with Polanyi's theory of the Great Transformation as economic development progresses.

Migration Migration of workers has been an important phenomenon over human history and migrants bring their motivation as well as skills. This has been a focus of an important body of work on culture in economics as developed and reviewed in Fernandez (2011) and Guiso et al (2006). This is now being based on a growing body of empirical evidence For example, Guiso et al (2004) show that migrants within Italy bring their levels of trust with them to their new locations. Another example is and Fernandez and Fogli (2009) which shows that second-generation American women appear to "inherit" patterns from a woman's ancestral country. Inspired by this work, our framework provides a way of thinking about the self-enforcing dynamics due to migration decisions in the context of worker motivation.

To model this formally, suppose that there is a pool of potential migrants of measure M and among the migrants, let μ^M be motivated. There are two economies A and B that the migrants are choosing to migrate to. These two economies are assumed to have the same structure of productivity and other parameters but A is assumed to have more motivated workers than B, i.e., $\mu^A > \mu^B$. To make migration worthwhile for all types of migrants, suppose there is an exogenous benefit α from migration which exceeds $-(1-N)\,\varepsilon$. This could, for example, be due to migrants fleeing persecution in their home country. Therefore, all of our potential migrants prefer moving to A or B rather than staying in their home country. We focus on the location decisions by the pool of motivated workers since all selfish workers will get a payoff of zero

in all economies (all else equal) and hence have strict preference between A or B and are equally likely to land in one or the other.[14]

Let $\Delta^A = \Delta\left(\mu^A\right)$ and $\Delta^B = \Delta\left(\mu^B\right)$ be the expected utility of a motivated worker in each economy. A motivated migrant will pick which economy to migrate to based on $\max\left\{\Delta^A, \Delta^B\right\}$.[15] Now we have:

Proposition 5 *Potential migrants will sort according to the fraction of motivated workers in each country. Specifically, if* $\Delta^A > \Delta^B$ $\left(\Delta^A < \Delta^B\right)$ *the fraction of motivated workers in A increases (decreases) to*

$$\frac{\mu^A N + \mu^M M}{N + M\mu^M + \frac{1}{2}\left(1 - \mu^M\right) M} \left(\frac{\mu^A N}{N + \frac{1}{2}\left(1 - \mu^M\right) M}\right).$$

Since there is positive complementarity from the fraction of motivated types, this will mean that a region or country which already has a larger fraction of motivated types will tend to attract more motivated types (and vice versa). When socialization is added to the change in the pool of workers due to migration, this effect will become even stronger as there is a dynamic cultural multiplier at work over time which can reinforce cultural differences.

Migration has the potential to change the trajectory of an economy if migration results in a movement of an economy to above or below $\hat{\mu}$. More generally, we would expect open labor markets and migration to foster cultural heterogeneity even when economic conditions are relatively similar.

6 Robustness

In this section, we discuss a few further issues that assess the robustness of the insights generated in the analysis above. We will argue that although we have focused on a specific case, the core insights and ideas remain relevant more generally.

[14] We are abstracting from the possibility the selfish workers anticipate the utility benefits from mixing with motivated workers in the destination country.

[15] As in the case of selfish workers, we assume that they are myopic and focus on one period benefits.

The Downside of Autonomy It is arguable that the model offers an excessively optimistic view of the power of intrinsic motivation. The psychological literature emphasizes that intrinsic motivation comes from autonomy. However, giving workers more autonomy that may mean that they can use their autonomy to deviate from profit maximizing strategies at the expense of firm owners. This would create a trade-off between allowing autonomy and incentivizing workers.

This idea can be captured simply by supposing that profits are only $\beta\pi$ when they employ motivated workers and do not have a monitoring and/or incentive schemes in place to counteract this. In this case, the critical level of profit where firms employ motivated workers becomes:

$$\hat{\pi} = \frac{2c}{1 - \beta\mu^2}.$$

This will lead to fewer opportunities for motivated workers and hence raise the critical threshold $\hat{\mu}$ above which society will converge to having only motivated workers in the population. In the long run, GDP will now be lower in economies which rely on intrinsic motivation to the extent that β is not a transfer.[16] Note that as $\beta \to 0$, then all firms will use incentives. The impact on welfare of having more intrinsic motivation now becomes ambiguous, i.e. there is a trade-off between a society based on intrinsic motivation being happier whereas one based on incentives will tend to have higher GDP all else equal.

Dynamic contracts The key insight from our basic model is the economy cannot immediately achieve separation between workers of different types and hence will tend to yield the kind of mixing where socialization is possible. This was facilitated by fresh recontracting between firms and workers in each period without memory. However, in some cases firms may be able to update about the types of their workers after having observed their performance. For

[16] Another reason might be that lower investment by capitalists since they now earn lower returns.

example a firm that does not use bonus pay would know, if profits are positive, that it had hired two intrinsically motivated workers and hence would wish to offer a contract to retain them in the next period. Observing that the firm is retaining the workers would, in turn, convey information to the market and low productivity firms would be susceptible to losing its intrinsically motivated workers to more productive firms.

In general, we would expect such processes to increase the extent of assortative matching at each point in time. However, only in extreme cases, would this yield complete separation and hence prevent the core socialization effect that we have focused on from breaking down. For example, workers would have to be infinitely lived and with no random terminations between firms and workers. However, in terms of the core model, it would lead σ in (6) to lie between zero and one and also to be time varying. Even if σ converges to zero (full assortative matching), the workplace socialization process would continue along the path to that point.

Monopsony We emphasized in our discussion of Proposition 1 that competition between firms plays a key role in the finding that only pooling contracts are possible. We believe that focusing on the case competitive labor allocation is a natural core case. We show in Appendix B that, if labor allocation were organized by a centralized social planner acting as a monopsonist, then it is possible to achieve separation of worker types when $\gamma \geq c$, i.e. the case that we have been studying. This shows the crucial role that competition plays in eliminating the possibility of separating equilibria as has been argued previously, for example, in the literature on adverse selection with insurance as in Rothschild and Stiglitz (1976).

The Anomie Assumption Finally consider what happens when $\gamma < c + \theta$. In this case, intrinsically motivated workers who work under incentive contracts are still happier than selfish workers. Now there is no anomie and over time, we would expect intrinsic motivation to expand and for $\mu = 1$ to be the unique steady state of the economy. This highlights a key role played

by (7) in the analysis and for the possibility that selfish workers (standard rational economic agents) come to dominate the landscape in the long run.

Our assumption in the main body of the paper seems reasonable. The core idea which anomie captures is the culture clash that occurs when practices within firms are at odds with the values of workers. If workers actually prefer to receive rewards in the form of incentive pay even when they are intrinsically motivated, then we would always expect the process of cultural evolution to support the propagation of intrinsic motivation which makes workers happier and more productive in all conditions.

This brief discussion suggests that a more realistic model would allow heterogeneity in γ and θ across individuals (possibly with shocks over time). What matters to the specific cultural dynamics brought out by the analysis above is that at least some workers suffer from the anomie which drives the cultural dynamics away from intrinsic motivation once there are sufficiently many firms who prefer to use bonus pay. The smaller is the set of workers who experience that sense of anomie, the more likely it is that a society will grow its level of intrinsic motivation over time, *ceteris paribus*.

7 Concluding Comments

Most economic studies of worker heterogeneity have focused on differences in ability often measured as cognitive skills. However, a range of recent research (such as Heckman 2007) have pressed the case for a wider appreciation of capacities which encompasses non-cognitive skills such as perseverance and motivation. This brings the psychological and economic perspectives on human differences close together. Just as human capital can enhance cognitive skills, there are ways of changing other human capacities. A range of influences are possible including parental investments and the influence of schools and peer groups which shape identities. Here, we have looked at workplace socialization and the role that this may have in influencing the cultural dynamics of motivation. The core arguments that we have developed would apply to a range of social interactions which shape motivation.

The approach to cultural evolution that we have emphasized here emphasizes the importance of encounters between individuals of different types. We have allowed the payoffs in the marketplace which influence the socialization process. Types which have better market opportunities are more likely to proliferate while those which do not would tend to contract. Thus, the model integrates market forces with an evolutionary approach based on a particular concept of payoff fitness.

The paper is unusual in taking a macro-economic perspective on the role of motivation. We look at how societies evolve in terms of this cultural characteristic. The framework that we have developed has a range of specific features in order to focus on some specific issues. Our approach has deliberately tried to take a minimalist approach to modelling to piece the argument together, consistent with the contracting frictions imposed by adverse selection and moral hazard.

In future work, it would be interesting to look at economies where there is a range of work-place cultures in different sectors that co-evolve. One particularly interesting issue is how far motivations which evolve in more public-service occupations differ from those that evolve in more profit-oriented sectors. It will also be interesting to explore other areas where intrinsic motivation is important in economic settings. This could include analyzing debates about the role of the market provoked by Titmus (1970) who emphasised the importance of differentiating between gift-based and market-based transactions as ways of exploiting and encouraging pro-social motivation.

References

[1] Akerlof, George, and Rachel Kranton, [2000], "Economics and Identity," *Quarterly Journal of Economics*, 115(3), 715–753

[2] Akerlof, George, and Rachel Kranton, [2010], *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton: Princeton University Press, 2010.

[3] Alesina, Alberto, and Paola Giuliano, [2015], "Culture and Institutions", *Journal of Economic Literature*, 53(4): 898-944.

[4] Besley, Timothy, [2015], "Aspirations and the Political Economy of Inequality," Hicks Lecture, University of Oxford.

[5] Besley, Timothy and Maitreesh Ghatak, [2005], "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3), 616-636.

[6] Benabou, Roland and Jean Tirole, [2003], "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70(3), 489-520.

[7] Benabou, Roland and Tirole, Jean, [2006], "Incentives and Pro-social Behavior," *American Economic Review*, 96(5), 1652-1678.

[8] Bisin, Alberto and Thierry Verdier, [2001], "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97, 298–319.

[9] Bisin, Alberto and Thierry Verdier, [2011], "The Economics of Cultural Transmission and Socialization," in Jess Benhabib (ed), *Handbook of Social Economics*, Volume 1A, Chapter 9, Elsevier.

[10] Boyd, Robert., and P. J. Richerson. [1985], *Culture and the Evolutionary Process*, Chicago: University of Chicago Press.

[11] Bramoullé, Y., H. Djebbari, and B. Fortinet, [2008], "Identification of peer effects through social networks," *Journal of Econometrics*, 150: 41-55

[12] Bryson, Alex, John Forth and Lucy Stokes, [2015], "Does Worker Wellbeing Affect Workplace Performance?" CEP Discussion Paper No. 1363.

[13] Cavalli-Sforza, L. L., and Feldman, M. W., [1981]. *Cultural Transmission and Evolution*, Princeton University Press, Princeton, N.J.

[14] Deci, Edward L. [1975], *Intrinsic Motivation*, New York: Platinum Press,

[15] Deci, Edward L. and Richard M. Ryan. [1985], *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

[16] Durkheim, Emile. [1893], *The Division of Labor in Society*, 1964 edition, New York: Free Press.

[17] Fernandez, Raquel and Alessandra Fogli, [2009], "Culture: An Empirical Investigation of Beliefs, Work and Fertility," *American Economic Journal: Macroeconomics* 1(1), 146–177.

[18] Fernandez, Raquel. [2011], "Does Culture Matter?," in, J. Benhabib, M. O. Jackson and A. Bisin, (eds), *Handbook of Social Economics*, Vol. 1A, North-Holland, 2011

[19] Festinger, Leon, [1957], *A Theory of Cognitive Dissonance*, California: Stanford University Press.

[20] Frey, Bruno, [1997], *Not Just for The Money: An Economic Theory of Motivation*, Cheltenham: Edward Elgar.

[21] Gelfand, M. et al., [2011], "Differences between tight and loose cultures: A 33-nation study," *Science*, 332: 1100-1104.

[22] Gittleman, Maury and Brooks Pierce, [2013], "How Prevalent is Performance-Related Pay in the United States? Current Incidence and Recent Trends," *National Institute Economic Review* 226: R4-R16.

[23] Gorodnichenko, Yuriy and Gerard Roland [2011], "Which Dimensions of Culture Matter for Long-Run Growth?" *American Economic Review* (Papers and Proceedings), 101(3): 492-498

[24] Guiso, Luigi, Paola Sapienza, and Luigi Zingales, L., [2004], "The Role of Social Capital in Financial Development," *American Economic Review*, 94, 526–556.

[25] Guiso, Luigi, Paola Sapienza, and Luigi Zingales, L., [2006], "Does Culture Affect Economic Outcomes?", *Journal of Economic Perspectives*, 20, 23–48.

[26] Güth,Werner, [1995], "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives," *International Journal of Game Theory*, 323–344.

[27] Güth,Werner and Menahem E. Yaari, [1992], "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach," in Ulrich Witt (ed.) *Explaining process and change: Approaches to evolutionary Economics,* Ann Arbor: University of Michigan Press, 1992, 23–34.

[28] Heckman, James J., [2007], "The technology and neuroscience of capacity formation," *Proceedings of the National Academy of Sciences*, 104(33), 13250–13255.

[29] Holmström, Bengt and Paul Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, & Organization*, Vol. 7, Special Issue, pp. 24-52.

[30] Kohn, Alfie, [1975], "Why Incentive Plans Cannot Work," *Harvard Business Review*, September-October, 16-21.

[31] Mas, Alexandre and Enrico Moretti, [2009] "Peers at Work," *American Economic Review*, 99 (1), 112–45.

[32] Merton, Robert K., [1938], "Social Structure and Anomie," *American Sociological Review*, 3, 672-682.

[33] Merton, Robert K. [1968], *Social Theory and Social Structure*, New York: Free Press.

[34] Ostrom, Elinor [2000], "Collective action and the evolution of social norms," *Journal of Economic Perspectives*, 10(3) 137-158.

[35] Polanyi, Karl [1944/2001], *The Great Transformation: The Political and Economic Origins of Our Time*, 2nd ed. Foreword by Joseph E. Stiglitz introduction by Fred Block. Boston: Beacon Press.

[36] Rothschild, Michael and Joseph Stiglitz, [1976], "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90(4), 629-649.

[37] Ryan, Richard M. and Edward L. Deci [2000], "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational Psychology* 25, 54–67.

[38] Sandel, Michael,[2012], *What Money Can't Buy : The Moral Limit of Markets*, London: Allen Lane.

[39] Schein, Edgar H. [1965], *Organizational Psychology*, Englewood Cliffs, NJ: Prentice-Hall

[40] Sen, Amartya K., [1977], "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6(4), 317-344.

[41] Sethi, Rajiv, and E. Somanathan, [2001], "Preference Evolution and Reciprocity," *Journal of Economic Theory*, 97, 273-297.

[42] Smith, Adam, [1976], *The Theory of Moral Sentiments*. Edited by D. D. Raphael and A. L. Macfie. Oxford: Clarendon Press; New York: Oxford University Press, (originally published in 1759).

[43] Tabellini, Guido, [2008], "The Scope of Cooperation: Norms and Incentives," *Quarterly Journal of Economics*, 123 (3), 905–950.

[44] Titmuss, Richard, [1970], *The Gift Relationship: From Human Blood to Social Policy*, London: Allen and Unwin.

[45] Van Maanen, John E., and Schein, Edgar H. [1979], "Toward a theory of organizational socialization," *Research in Organizational Behavior* 1: 209-269.

[46] Weber, Max, [1922], *Economic and Society*, University of California Press, 1978 (originally published in 1922).

Appendix

A Proofs

Proof of Lemma 1: Let $\{i, j\}$ be a pair of agents in a given firm where the wage contract has a bonus component b. If both i and j are of type s then we look for a Nash equilibrium where:

$$\begin{aligned} e_i &= \arg \max_{e \in \{0,1\}} \{e\,[e_j b - c]\} \\ &\quad \text{and} \\ e_j &= \arg \max_{e \in \{0,1\}} \{e\,[e_i b - c]\}\,. \end{aligned}$$

For all b, there is a Nash equilibrium where $e_i = e_j = 0$. For $b \geq c$, it is straightforward to check that there is also a Nash equilibrium where $e_i = e_j = 1$ which payoff dominates the equilibrium with $e_i = e_j = 0$ for $b \geq c$, and strictly so for $b > c$. For $b < c$, even if the other worker chooses $e_j = 1$, worker i will choose $e_i = 0$. Therefore the unique equilibrium is $e_i = e_j = 0$ for $b < c$.

When a selfish agent i is paired with a motivated agent j, his effort decision, conditional on the motivated worker putting in effort is

$$e_i = \arg \max_{e \in \{0,1\}} \{e\,[b - c]\}\,.$$

Therefore, given the tie-breaking assumption when $b = c$, it follows directly that $e_i = 1$ if and only if $b \geq c$. ■

Proof of Proposition 1: **Step 1**: *The allocation $e_m = e_s = 1$ is an equilibrium allocation if and only if the contracts have the property $b_s \geq c$ and $b_m \geq c$.*

If $b_s \geq c$ then selfish workers would (weakly prefer to) choose a positive effort level, while motivated workers always do, so that necessity is straightforward. To prove sufficiency, let $e_m = e_s = 1$. From Lemma 1, we know that for this to be an equilibrium, a necessary condition is $b_s \geq c$. Therefore, we

proceed to prove $b_m \geq c$. The self-selection constraint of the selfish type is

$$b_s - c + w_s \geq V^s(b_m) + w_m.$$

The competition constraint is

$$w_m + b_m = w_s + b_s.$$

Substituting

$$-c \geq V^s(b_m) - b_m.$$

If $b_m = 0$ then $V^s(b_m) = 0$ and $-c \geq 0$ cannot hold. If $c > b_m > 0$, then $V^s(b_m) = b_m$ and once again the (2) cannot hold. However, if $b_m \geq c$ then $V^s(b_m) = b_m - c$ (as selfish types put in effort) and so the condition holds. Now we check the (2) holds for the motivated workers:

$$w_m + V^m(b_m) \geq w_s + V^m(b_s).$$

For $b_m \geq c$ and $b_s \geq c$

$$w_m + b_m - \gamma + \theta \geq w_s + b_s - \gamma + \theta$$

which holds with equality given the competition constraint.■

Step 2: *If* $e_m = e_s = 1$ *then the unique competitive equilibrium is* $b_s = b_m = c$ *and* $w_s = w_m = z$.

The competition constraint implies $b_s + w_s = b_m + w_m$ and firms choose the minimum wage and bonus payments subject to the minimum consumption and limited liability constraints (3). Since $e_m = e_s = 1$, we know both b_m and b_s would have to be at least as large as c. Also, the minimum consumption constraint implies both w_s and w_m would have to be at least as large as z. For there be a separating equilibrium with $b_s \neq b_m$ and $w_s \neq w_m$ it must be

$b_i > c = b_j$ and $w_j > z = w_i$ since having both the bonus exceed c and the flat wage exceed z is not consistent with profit-maximization when the number of workers exceed the number of jobs. But by extending the same logic, a firm could introduce a contract $b' = c$ and $w' = z$ that respects all the above constraints and makes more profits. The participation constraint is:

$$\min\{c + z - c, \theta + c + z - \gamma\} = \min\{z, \theta + c + z - \gamma\} \geq z - \varepsilon.$$

It is clearly satisfied for selfish workers, and for motivated workers it is satisfied if $\theta - \gamma + c \geq -\varepsilon$, which is what we assume. Therefore, such a contract will be able to attract workers of both types and make strictly more profits than the pair (b_i, w_i) and (b_j, w_j).■

Step 3. *Consider the allocation* $e_s = 0$ *and* $e_m = 1$. *A necessary and sufficient condition for this to be part of an equilibrium allocation is* $b_m = b_s = 0$ and $w_m = w_s = z$.

First, consider necessity and let $b_m = b_s = 0$ and $w_s = w_m = z$. The type s workers will not work while the type m workers set $e_m = 1$. Since it is a unique contract, self-selection is not relevant. The payoff of motivated workers is $z + \theta$ while that of selfish workers is z so both choose to participate given that $\varepsilon > 0$ and so being employed is strictly preferred to being unemployed. This contract is therefore able to elicit $e_s = 0$ *and* $e_m = 1$.

To prove sufficiency, suppose now that $e_s = 0$ and $e_m = 1$. The payoff to a firm who employs two motivated workers is

$$\pi - 2(b_m + w_m).$$

No output is produced in firms that employ one selfish and one motivated worker or two selfish workers, and their payoffs are $-2w_s$ or $-(w_s + w_m)$. These three cases occur with probabilities $\mu^2, 2\mu(1-\mu)$, and $(1-\mu)^2$ respectively. Maximizing expected profits subject to (3) yields $b_m = 0$ and $w_m = w_s = z$. Since $e_s = 0$, $b_s < c$. Since b_s is not actually paid, the only

reason to possibly choose $b_s > 0$ would be to relax the self-selection constraints:

$$\begin{aligned} z + \theta &\geq b_s + z + \theta - \gamma \\ &\text{and} \\ b_s + z &\geq z. \end{aligned}$$

Since $b_s = 0$ satisfies the self-selection constraints, in equilibrium $b_s = 0$ since $b_s > 0$ involves higher cost to firms with no benefits. The contract $b = 0$ and $w = z$ satisfies the voluntary participation constraint of the motivated worker whose payoff is $z + \theta$, as well as that of the selfish worker whose payoff is z as both these quantities are higher than $z - \varepsilon$.■

Step 4. Under the first type of pooling equilibrium ($e_m = e_s = 1$), a firm's profit is

$$\pi - 2(z + c).$$

Under the second type ($e_s = 0$ *and* $e_m = 1$), a firm's expected profit is:

$$\mu^2 \pi - 2z.$$

Therefore, the positive bonus contract will be chosen by a firm when

$$\pi - 2(z + c) > \mu^2 \pi - 2z$$

or,

$$\pi > \frac{2c}{1 - \mu^2}.$$

Under both contracts, selfish workers have the same payoff, namely, z. However, motivated workers earn $z + c + \theta - \gamma$ under the first type of contract and $z + \theta$ under the second type of contract. Under our assumption, they are better off in the latter case. Firms are indifferent between selfish and motivated workers under the first type of contract, while in the second type of contract

they strictly prefer motivated workers. Since the alternative to not finding a job is to engage in the subsistence activity that yields a payoff of $z - \varepsilon$, given the assumption of random allocation of workers seeking employment to firms, motivated workers apply to both types of firms. ■

Proof of Proposition 2: Observe that

$$\mu_{t+1} - \mu_t = N\mu_t (1 - \mu_t) \left[2\rho \left(\left[\theta + \left(1 - G \left(\frac{2c}{(1 - \mu_t^2)} \right) \right) (c - \gamma) \right] \right) - 1 \right].$$

Now it is clear that for $\mu_t > \hat{\mu}_t$ then $\mu_{t+1} - \mu_t > 0$ and the fraction of motivated types grows over time. Moreover, if $\Delta(\mu_t) > 0$, then $\Delta(\mu) > 0$ for all $\mu \geq \mu_t$. Then there is a steady state where $\mu = 1$. A similar argument covers the case where $\mu_t < \hat{\mu}$ where μ converges to $\mu = 0$. ■

Proof of Proposition 4: Let

$$\theta = \left(1 - G^J \left(\frac{2c}{\left(1 - \left(\hat{\mu}^J \right)^2 \right)} \right) \right) [\gamma - c] \text{ for } J \in \{A, B\}.$$

Then

$$G^A \left(\frac{2c}{\left(1 - \left(\hat{\mu}^A \right)^2 \right)} \right) = G^B \left(\frac{2c}{\left(1 - \left(\hat{\mu}^B \right)^2 \right)} \right)$$

if and only if

$$\frac{2c}{\left(1 - \left(\hat{\mu}^A \right)^2 \right)} \geq \frac{2c}{\left(1 - \left(\hat{\mu}^B \right)^2 \right)}$$

or $\hat{\mu}^A \geq \hat{\mu}^B$.■

B A Monopsonistic Labour Market

The monopsonist must respect the incentive-compatibility conditions.

Suppose that $b_s \geq c$ and $b_m = 0$. Then the two truth-telling constraints

are:

$$b_s - c + w_s \geq w_m \text{ and } w_m + \theta \geq w_s + b_s + \theta - \gamma.$$

Combining these yields

$$b_s - c \geq w_m - w_s \geq b_s - \gamma.$$

This cannot hold if $\gamma < c$ since $b_s - c \geq 0$ and if $\gamma < c$, then $b_s - \gamma > b_s - c$.

Suppose instead that $b_s \geq c$ and $b_m \geq c$. The two truth-telling constraints are now

$$\begin{aligned} b_s - c + w_s &\geq w_m + b_m - c \\ w_m + b_m + \theta - \gamma &\geq w_s + b_s + \theta - \gamma. \end{aligned}$$

Combining these yields

$$b_s - b_m \geq w_m - w_s \geq b_s - b_m$$

which can only hold if

$$b_s - b_m = w_m - w_s$$

which is consistent with competition. Hence we have shown that if $\gamma \geq c$, then there is a separating equilibrium under monopsony.