The Evolution of Motivation[*]

Timothy Besley   Maitreesh Ghatak

LSE and CIFAR   LSE

January 2023

Abstract

This paper studies evolution of motivation when firms workers care about the mission of the firm. New workers who join an organization are socialized by those who work there already. We show that there is a natural complementarity between choosing a mission to suit worker motivation and having a large stock of motivated agents. The model has multiple steady states and we discuss economic factors which shape the emergence of mission motivation as a cultural trait.

JEL Classification: L23, M14, M52

Keywords: worker motivation, incentives, dynamics of preferences

1 Introduction

It is now routine to question the narrow view of human motivation caricatured by the idea that *homo economicus* is a rational egoist (see, for example Fehr and Falk, 2002). A range of important insights from psychology and experimental evidence have opened the black box of human motivation as an object of study in economics. Studying whether and how motivation changes over time has, however, received less attention. But a widely accepted idea in organizational psychology is that agents are socialized in the work place and that their motivations, values and preferences are therefore endogenous.

This paper explores the dynamics of workplace motivation from a cultural-evolutionary perspective emphasizing the interplay between rewards structures and the psychological fitness of different motivational types creating dynamics of motivation in organizations. The key question we ask is, if some agents are driven partly by non-pecuniary motivation, while others have standard preferences of selfish economic agents, can the former type survive in the long-run given how a profit-maximizing firm will use incentive schemes anticipating a certain distribution of types of agents, and the distribution of types evolves based on fitness advantage according to agents' payoffs?

The paper focuses on mission motivation of the kind emphasised in Besley and Ghatak (2005) who propose that agents are willing to put in effort in firms which produce output in a particular way such as respecting environmental goals or treating their clients better. They focused on the role of matching between firms and workers to take advantage of mission preferences. Here, we suppose that matching is random and firms are profit-maximizing. But motivation is endogenous due to socialization of workers within a firm.

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. Our ideas also relate to the historical sociological literature such as Durkheim (1893) and Polanyi (1944) who saw changes in the nature of the employment relationship as one the central cultural processes which evolved with the advent of a market economy.

The key message of the paper is that movements away from the *homo economicus* assumption make sense when there are good reasons to believe that alternative motivations have greater psychological fitness according to specific criteria. In organizational settings where particular kinds of motivations thrive, we would expect to see them grow according to any reasonable model of cultural evolution. But that depends on how organizations treat their motivated agents in pursuit of organizational objectives where our baseline case is profit maximization. We show precisely when profit maximization is consistent with a process of cultural evolution which yields mission motivation in the long-run. But equally, the model shows when such motivation is fragile.

The remainder of the paper is organized as follows. The next section discusses some related literature. In section three, we lay out the approach. Section four develops implications of the ideas. Section five offers some concluding remarks.

2   Related Literature

Mission motivation   This paper is related to range of approaches to a more psychologically informed theory of human motivation as discussed, for example, in Lazear (1991) and Kamenica (2012). The specific approach that we take follows Besley and Ghatak (2005) who suppose that workers can be motivated by non-pecuniary "mission" goals. There are many examples – doctors who are committed to saving lives, researchers to advancing knowledge, judges to promoting justice and soldiers to defending their country in battle. Viewing workers as mission-oriented makes sense when the output of the mission-oriented sector is thought of as producing collective goods. The benefits and costs generated by mission-oriented production organizations are not fully reflected in the market price. In addition, donating our income earned in the market to an organization that pursues a mission that we care about is likely to be an imperfect substitute to joining and working in it. This could be due to the presence of agency costs or because individuals care not just about the levels of these collective goods, but their personal involvement in their production (i.e., a "warm glow"). In this paper we focused on the interaction between the selection of workers across firms in terms of their motivation and what incentive contracts are offered by firms (e.g. firms that offer high-powered monetary incentives may attract extrinsically motivated

3

types). In the current paper we focus on how the motivation of workers evolve within a firm.

This 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).

Mission motivation is a particular form of intrinsic motivation which has been widely discussed in the psychology literature on motivation.[1] 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. Benabou and Tirole (2003, 2006) 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 how 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,

---

[1]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 purely 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. 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.

one-shot interactions.

Carpenter and Gong (2016) uses a lab experiment to confirm that motivated workers will produce higher output, and financial incentives can largely substitute for mission motivation when workers and employers are mismatched in mission preferences.

Hiller and Verdier (2014) explores how market structure affects firms' investment in corporate culture, i.e. the cultural homogeneity that align workers to firm's objectives and can help to substitute monetary incentives: a larger product market size and higher competition for managers on the labor market induces firms to invest more in corporate culture and reduce financial incentives.

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).[2] The model developed here shares the core structure of population dynamics with this approach. 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) which has been explored in detail in Alger and Weibull (2013), Dekel et al (2007) and Sethi and Somanathan (2001).

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 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.[3] Bidner and Francois (2012) develop a general equilibrium where norm-driven behavior evolves endogenously.

---

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

[3]See Bisin and Verdier (2011) for an overview of this literature. Bisin and Verdier (forthcoming) provides the most recent review on cultural transmission, which discusses the basic framework, the micro-foundations of the cultural transmission technology, as well as how it can be extended to explore the interactions among culture, social economic-environments, and institutions.

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. In this spirit, Francois and Zabojnik (2005) study how trust norms evolve in the process of economic development. Here we focus on the complementarity between reward structures and the type-space as well as looking at alternative sources of non-pecuniary motivation.

In related earlier work (Besley and Ghatak, 2016) we have explored the role of competition among firms when the type of workers is subject to asymmetric information and the co-evolution of the reward structure in a competitive labor market and the distribution of motivation in the workforce. In that paper, the focus is on the competition among firms for workers whose types are subject to asymmetric information, and how the consequent evolution of motivation affects overall productivity. These are the key distinguishing features relative to the present exercise, where the focus is on alternative foundations of non-pecuniary motivation within a single firm context, with no asymmetric information.

Dessein and Prat (2022) studies the evolution of organizational capital, which they define as firm's intangible asset that slowly changes over time, and produced by firm's CEOs. This is similar to our notion of mission motivation where motivated workers tend to produce higher outputs; but we emphasize the evolution of mission motivation as a result of socialization among workers instead of the CEO's choice.

Besley and Persson (2019) studies the two-way interaction between democratic values and instituitions, where some citizens hold values that make them protest to preserve democracy with the share of such citizens evolving endogenously over time. The papers shows that there is a natural complementarity between values and institutions creating persistence.

Besley and Persson (2022) study the interaction between the organization design in terms of degree of centralization, and organizational culture, where culture is modeled as social identity in the workplace that affect project choices. They study cultural dynamics depending on the expected relative payoffs from holding different identities, and study the role of delegation when the alignment between culture and organizational goals vary.

3 The Model

We consider a firm that operates in isolation of others. We can think of this as a labor market where firms have very high levels of specific human capital and individuals join a firm for life. Turnover is then purely due to death or illness. In this world, the outside option of an existing worker is not to work for the firm and engage in an activity that yields an expected return normalized at 0.[4]

Following Besley and Ghatak (2005) we will allow for the possibility of there being mission-motivated workers. An organization comprises a continuum of agents indexed $i \in [0,1]$ each of which is one of two types $\tau \in \{m, s\}$ where m stands for "motivated" and s standards for "selfish". We will provide an explicit formulation of how these two types of workers are differentiated below. Let $\mu \in [0,1]$ be the fraction of motivated workers.

Production Technology   Each agent put in a unit of effort $e \in \{0,1\}$. Let individual output be $x(e)$ where $x(0) = 0$ for all $i \in [0,1]$ and

$$x(1) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p. \end{cases}$$

Therefore, contingent on choosing $e = 1$, the expected output of an agent is p.[5] We assume that the output realization of workers are independent. Let λ be the proportion of agents in the organization who set $e = 1$.[6]   Expected *total* output of the firm then is $X(\lambda) = \lambda p$. The firm earns a revenue of y per unit of output net of costs of non-labour inputs.

Information and Agent Motivation   The effort choices of workers e are unobservable to the firm, but the output $x(e)$ is verifiable.   Both types of

---

[4]In a related paper (Besley and Ghatak, 2016) we explore the role of competition among firms for workers in a setting similar to this.

[5]It would be straightforward to introduce the possibility that a worker can produce some baseline output even with low effort without affecting any of the main results.

[6]As there is a continuum of workers, formally we have

$$\lambda = \left(\int_0^1 e(i)\, di \right).$$

Following Judd (1985), we assume that this integral is well-defined. Uhlig (1996) and Al-Najjar (2004) show that this will be the case if we use a Pettis integral.

workers incur a disutility of effort from $e = 1$ denoted by ψ. We assume that $\psi \in [0, \bar{\psi}]$ with distribution function $F(\psi)$ and each worker receives an idiosyncratic draw from the distribution each period, with associated strictly positive density $f(\cdot)$.

Motivated agents care about the mission of the firm. We suppose that creating such motivation is costly to the firm and takes the form of an initial investment each period before contracts to workers are offered and production takes place. We model this as choosing a discrete action $\sigma \in \{0, 1\}$ and the cost of mission choice is given by $c\sigma$. This could be a pro-social action, such as a cosmetics firm which does not use animal testing or using a low-carbon technology. It could also be a private action which generates a local public good for the workers such as free coffee or social events.

Conditional on choosing $e = 1$, motivated and selfish workers receive a potential non-pecuniary payoff $v_\tau(\sigma)$, $\tau \in \{s, m\}$ that partly offsets the disutility of effort. For selfish agents, this payoff is zero irrespective of the choice of σ. For motivated agents, however, it is influenced by the action taken by the firm. The firm does not observe worker type and offers the same contract to all workers.

We assume a utility function of agents that is linear in consumption and cost of effort:

$$U^\tau(z, e) \equiv z + e\left[v_\tau(\sigma) - \psi\right], \quad \tau \in \{s, m\} \tag{1}$$

where z is private consumption.

When they choose $e = 1$, motivated agents get a non-pecuniary payoff of $\theta > 0$ if the firm chooses $\sigma = 1$. However, conditional on choosing $e = 1$, they also get disutility $\varepsilon > 0$ (which can be arbitrarily small) if the firm does not have any mission and is just interested in maximizing profits. This is captured with the following specification:

$$v_m(\sigma) = \begin{cases} \theta & \sigma = 1 \\ -\varepsilon & \text{otherwise.} \end{cases}$$

The assumption that motivated agents earn some disutility when the firm has no mission matters for the dynamics of motivation, as it creates some potential payoff disadvantage for their type.

For selfish agents, $v_s(\sigma) = 0$ for $\sigma \in \{0, 1\}$.

Optimal Effort Since output for each worker takes two values, 0 and 1, and there is no correlation in output realization across workers, without loss

of generality we can focus on wage contracts that have a flat wage component $\omega \geq 0$ and a bonus component $\beta \in [0, y]$ for high output.

Optimal choice of effort for an agent with disutility draw ψ accordingly solves:

$$e\left(\beta p + v_\tau\left(\sigma\right) - \psi\right) = \arg \max_{e \in \{0,1\}} \left\{e\left[\beta p + v_\tau\left(\sigma\right) - \psi\right]\right\}.$$

This defines a cutoff level for ψ below which an agent chooses $e = 1$ defined by

$$\hat{\psi}_\tau\left(\beta, \sigma\right) = \left\{ \begin{array}{cc} \beta p + v_\tau\left(\sigma\right) & \tau = m \\ \beta p & \tau = s. \end{array} \right.$$

Average effort is given by:

$$\hat{\lambda}\left(\sigma, \mu, \beta\right) \equiv \mu F\left(\beta p + v_m\left(\sigma\right)\right) + \left(1 - \mu\right) F\left(\beta p\right). \tag{2}$$

Expected output is therefore $\hat{\lambda}\left(\sigma, \mu, \beta\right) p$. Notice that $\hat{\lambda}\left(\sigma, \mu, \beta\right)$ is always increasing in β as we would expect. Higher output can also be achieved by setting $\sigma = 1$ via its impact on motivated agents. For $\sigma = 1$, $v_m\left(\sigma\right) = \theta$ and $\hat{\lambda}$ is increasing in μ and θ. Otherwise, $\hat{\lambda}$ is decreasing in μ.

Contracts   Firms are profit maximizing and choose optimal labor contracts and mission: $\{\beta, \omega, \sigma\}$ Profits per worker are

$$\left(y - \beta\right) ep - \omega.$$

The expected total profit of the firm, aggregating over all workers, is:

$$\left[y - \beta\right] \lambda p - \sigma c - \omega.$$

Since the firm's choice of σ is made before all workers are offered contracts we solve the firm's problem in two steps.   First, we choose the incentive scheme $\left\{\hat{\omega}_\sigma\left(\mu\right), \hat{\beta}_\sigma\left(\mu\right)\right\}$ for given σ. We then determine σ optimally.

Incentives are determined by profit maximization:

$$\hat{\Pi}\left(\mu, \sigma\right) = \max_{\{\omega, \beta\}} \left\{\left[y - \beta\right] \hat{\lambda}\left(\sigma, \mu, \beta\right) p\right\} - \omega.$$

It is clear that profits are decreasing in ω hence it is optimal to set the fixed wage as low as as possible. Henceforth, we will therefore set $\omega_\sigma\left(\mu\right) = 0$ for

9

$\sigma \in \{0,1\}$ and $\mu \in [0,1]$. Hence, the only optimization decision for firms is over β. The first-order condition for the choice of β is

$$-\hat{\lambda}(\sigma,\mu,\beta) + [y - \beta] \frac{\partial \hat{\lambda}(\sigma,\mu,\beta)}{\partial \beta} = 0.$$

This can be rewritten as:

$$\frac{1}{y - \beta_\sigma(\mu)} = \frac{1}{\hat{\lambda}(\sigma,\mu,\beta)} \frac{\partial \hat{\lambda}(\sigma,\mu,\beta)}{\partial \beta}. \tag{3}$$

We have noted before that $\hat{\lambda}$ is increasing in β and given that β cannot exceed y, (3) enables us to find an interior solution for β.

The optimal decision of the firm regarding the choice of β involves balancing the marginal cost of providing incentives in terms of lower net profits against the incentive benefits from rewarding agents more for high output. Using (2), the first-order condition can be rewritten as:

$$\frac{1}{y - \beta_\sigma(\mu)} = \frac{\partial \log \left(\mu F(\beta p + v_m(\sigma)) + (1 - \mu) F(\beta p) \right)}{\partial \beta}.$$

We make the following regularity assumption concerning the distribution function $F(.)$ which covers all of the applications below:

Assumption 1: We make the following assumption: *(i) $F(\psi)$ is a log concave distribution; (ii) $\lambda = \mu F(\beta p + v) + (1 - \mu) F(\beta p)$ is log concave in β ; (iii) $F'(0)$ is bounded; (iv) $\theta < h^{-1}(\frac{1}{py})$, where $h(\psi) = \frac{F'(\psi)}{F(\psi)}$.*

This assumption holds for standard distributions like uniform, exponential, and Pareto. This assumption directly implies the following result regarding $\beta_0(\mu)$ and $\beta_1(\mu)$, i.e., the optimal choice of β when $\sigma = 0$ and 1 respectively:

Lemma 1: *Suppose Assumption 1 holds. Then $\beta_0(\mu), \beta_1(\mu) > 0$ for all $\mu \in [0,1]$ and $\beta_0(\mu)$ is increasing in μ. Also, $\beta_1(\mu)$ is decreasing in μ and θ.*

Assumption 1 implies that the marginal effect of bonuses on total effort is lower : (i) when the non-pecuniary payoff that motivated agents receive is higher; (ii) when there are more motivated agents in the population; (iii) the

higher the level of the bonus (which ensures that the second-order condition is satisfied).

The profit of the firm with $\sigma = 0$ is

$$\hat{\Pi}(\mu, 0) = \max_{\beta \geq 0} \left\{ (y - \beta) \, \hat{\lambda}(0, \mu, \beta) \right\}.$$

For $\sigma = 0$, we know from Assumption 1 that $\hat{\beta}_0(\mu) > 0$. With $\sigma = 1$ is

$$\hat{\Pi}(\mu, 1) = \max_{\beta \geq 0} \left\{ [y - \beta] \, \hat{\lambda}(1, \mu, \beta) - c \right\}.$$

Now we ready to state our first result:

Proposition 1 *When the firm chooses a pro-social mission ($\sigma = 1$) incentives are flatter, i.e., $0 < \hat{\beta}_1(\mu) < \hat{\beta}_0(\mu)$.*
Proof: Note that from Lemma 1 we have shown that $\beta_1(\mu)$ is decreasing in μ and $\beta_0(\mu)$ is increasing in μ. Hence we know

$$\begin{aligned} \beta_0(0) \leq \beta_0(\mu) &\leq \beta_0(1) \\ \beta_1(1) \leq \beta_1(\mu) &\leq \beta_1(0). \end{aligned}$$

Observe that $\beta_0(0) = \beta_1(0)$. To see this, from the first order condition, we take $\mu = 0$ and we know that $\beta_0(0)$ and $\beta_1(0)$ are characterized by the same equation:

$$\frac{1}{y - \beta_\sigma(\mu)} = \frac{\partial \log(F(p\beta_\sigma(\mu)))}{\partial \beta}.$$

Therefore we have $\beta_1(\mu) \leq \beta_1(0) = \beta_0(0) \leq \beta_0(\mu)$.∎

In other words, agent motivation and financial incentives are substitutes and so using bonuses is less attractive, all else equal, in a world of motivated agents, paralleling one of the main results in Besley and Ghatak (2005), although here this is due to a deliberate and profit-maximizing choice of mission by the firm.

Socialization and Dynamics of Agent Type   Turning to the dynamics now, we assume that a type (of an agent) receives a fitness advantage based

11

on the average income for that type. For an agent of type τ, the expected payoff is:

$$Y^{\tau}(\mu) = F\left(p\beta_{\hat{\sigma}(\mu)}(\mu) + v_{\tau}(\hat{\sigma}(\mu))\right) p\beta_{\hat{\sigma}(\mu)}(\mu), \ \tau = m, s.$$

The material fitness advantage of the motivated type is defined as:

$$\Delta(\mu) = Y^m(\mu) - Y^s(\mu).$$

This function is critical to the evolution of motivation and we derive it for each application that we develop below.[7]

We assume that the evolutionary dynamics is "Darwinian" in the sense that the increase in the proportion of motivated agents is driven by their fitness advantage.[8] For our analysis we assume that there is a well-behaved function $Q(\mu, \Delta)$ that is *increasing* in Δ (and can depend on μ in a number of ways including being independent of it) :

$$\mu_{t+1} - \mu_t = Q(\mu_t, \Delta(\mu_t)). \tag{4}$$

We will work with a specific formulation which delivers such dynamics.[9]

Suppose that there is turnover in the organization each period with a fraction ρ of the workers who are replaced each period. *All* newly hired agents are assumed to be selfish but can be socialized on arrival by being mentored by an existing worker chosen at random. If she is mentored by a motivated agent, which happens with probability μ_t, we assume that she may become motivated depending on the relative psychological fitness of motivated and selfish types. In other words, socialization is based on material rewards received by the two types of agents. Moreover, we assume that this is something that can be observed by workers within an organization and it is also comparable across types, i.e. money has similar worth to both motivated and selfish types.

A randomly selected new agent is matched with an existing agent who is motivated with probability μ_t. Such an agent becomes motivated through mentoring by a motivated agent if:

$$\Delta(\mu_t) + \eta \geq 0,$$

---

[7] An explicit expression for $\Delta(\mu)$ is given in the proof of Propositions 2 below.

[8] From a philosophical point of view, this assumes that it is possible to compare utility across types for the same individual. In particular, we assume that an individual can figure out what their utility would be, if they were of a different type.

[9] Note that we have μ_{t+1} depending on μ_t and so this is a form of adaptive expectations where the fitness is measured for the contemporaneous value of μ_t.

where η is a mean-zero, *symmetrically* distributed idiosyncratic shock with a continuous distribution function $G(\cdot)$. We assume that $G(0) = \frac{1}{2}$, $G(\Delta(\mu_t)) > \frac{1}{2}$ for $\Delta(\mu_t) > 0$, and $G'(\Delta(\mu_t)) > 0$ for all $\Delta(\mu_t)$.[10]

Let $g(\cdot)$ be the density function corresponding to $G(\cdot)$. The probability that a new recruit mentored by a motivated type becomes motivated is the probability that $\eta \geq -\Delta(\mu_t)$, which is $1 - G(-\Delta(\mu_t))$. Given the symmetry assumption, this is equal to $G(\Delta(\mu_t))$.

Despite being matched with an existing agent who is motivated, if such direct socialization fails, the new recruit may still be indirectly socialized by observing and learning from other workers. The probability of indirectly becoming a motivated type depends monotonically on the average fraction of such types in the organization, a kind of social learning postulated in much of the cultural-evolution literature (Bisin and Verdier, 2001). Assuming a linear relation, the probability of indirect socialization becomes $(1 - G(\Delta(\mu_t)))\mu_t$ where μ_t is the fraction of motivated agents in the existing workforce at the beginning of period and $1 - G(\Delta(\mu_t))$ is the fraction of new agents for whom $\eta < -\Delta(\mu_t)$.

Adding these expressions, the overall probability that a new recruit who is matched with a motivated agent becomes motivated is:

$$G(\Delta(\mu_t)) + \{1 - G(\Delta(\mu_t))\}\mu_t. \tag{5}$$

If a new worker is matched with and mentored by a selfish worker instead, which happens with probability $1 - \mu_t$, there are two possibilities. First, she can be socialized into being selfish if

$$\Delta(\mu_t) + \eta \leq 0.$$

Thus, $G(-\Delta(\mu_t)) = 1 - G(\Delta(\mu_t))$ is the proportion of selfish workers coming from such matches. Second, she can indirectly become motivated (as above) depending on the aggregate fraction of motivated agents (μ_t) in the organization. The resulting probability of becoming motivated is therefore:

$$G(\Delta(\mu_t))\mu_t. \tag{6}$$

---

[10]An example would be the logistic distribution where the probability of a randomly selected new agent to become motivated through mentoring is:

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

It is easy to verify that the listed properties are satisfied.

The probability of a new agent being matched with a motivated or selfish agent being μ_t and $1-\mu_t$, multiplying (5) by μ_t and (6) by $1-\mu_t$, and adding, we get the fraction of new agents who become motivated agents. The overall fraction of motivated agents in the next period, μ_{t+1}, is therefore:

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

Simplifying the resulting expressions yields the following the equation of motion for the share of motivated types:

$$\mu_{t+1} = \mu_t + \rho\mu_t\,(1-\mu_t)\,[2G\,(\Delta\,(\mu_t)) - 1]. \qquad (7)$$

From this it is clear that studying the evolutionary dynamics of motivation requires studying the properties of $\Delta\,(\mu_t)$, in particular, its sign and how it changes with respect to μ_t.

Timing   The timing of the model is as follows:

1. At the beginning of each period, an organization inherits a fraction of motivated workers μ_t.

2. The firm chooses organizational form $\sigma \in \{0,1\}$

3. Workers are offered contracts

4. Agents choose their effort level

5. Output and payoffs are realized.

6. A fraction ρ of workers are replaced and new workers are socialized.

Steady States   A steady state, denoted by μ^*, requires three conditions to hold simultaneously relating to organizational form (σ), expected total output (X), and :

(i) $\sigma^* = \hat{\sigma}\,(\mu^*)$, (ii) $X^* = X\left(\hat{\lambda}\,(\sigma^*,\mu^*)\right)$, and (iii) $\mu^*\,(1-\mu^*)\,[2G\,(\Delta\,(\mu^*)) - 1] = 0$.

It is immediate from (iii) above that there is always a steady state where $\mu^* = 0$ and one where $\mu^* = 1$.

We focus only on *stable* steady states defined as follows. Recall the evolution of motivation satisfies

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

To explore stability, we examine

$$\frac{\partial\mu_{t+1}}{\partial\mu_t} = 1 + \rho(1 - 2\mu_t)[2G(\Delta(\mu_t)) - 1] + 2\rho\mu_t(1 - \mu_t)g(\Delta(\mu_t))\Delta'(\mu_t).$$

Evaluating the partial derivative respectively at $\mu_t = 0$ we have:

$$\left.\frac{\partial\mu_{t+1}}{\partial\mu_t}\right|_{\mu_t=0} = 1 + \rho[2G(\Delta(0)) - 1] = 1 - \rho + 2\rho G(\Delta(0)) < 1,$$

since $\Delta(0) < 0$ and $G(\Delta(0)) < \frac{1}{2}$. And at $\mu_t = 1$, we have:

$$\left.\frac{\partial\mu_{t+1}}{\partial\mu_t}\right|_{\mu_t=1} = 1 - \rho[2G(\Delta(1)) - 1] < 1,$$

since $\Delta(1) > 0$ and $G(\Delta(1)) > \frac{1}{2}$.

In principle, there could also be a steady state where $\Delta(\mu^*) = 0$, since given our assumption of symmetry, $G(0) = \frac{1}{2}$. However, as we show below, this does not arise in our setting. The reason is, as we see below from the structure of $\Delta(\mu_t)$, $\Delta(\mu_t) \neq 0$ in any static equilibrium, and therefore, we know $2G(\Delta(\mu_t)) - 1 \neq 0$ for any μ_t. Hence we restrict attention to the two possible steady states, i.e. $\mu = 0$ and $\mu = 1$, both of which are stable.

4   Implications

Choice of Mission   Now we turn to the choice of σ by the firm. The mission choice of the firm solves:

$$\hat{\sigma}(\mu) = \max_{\sigma \in \{0,1\}} \hat{\Pi}(\mu, \sigma) - c\sigma.$$

Hence the firm will only choose to create an environment which motivates workers when there is a sufficient increase in profit from doing so.

Proposition 2 *Expected output $\hat{\lambda}(\sigma, \mu, \beta_1(\sigma))$ is higher under $\sigma = 1$ compared to $\sigma = 0$ when incentives are set optimally.*

15

Proof: To see this note that for any level of μ, when $\sigma = 1$ is chosen, it must yield higher expected profits to the firm:

$$(y - \beta_1(\mu))\,\hat{\lambda}(1, \mu, \beta_1(\mu)) - c - (y - \beta_0(\mu))\,\hat{\lambda}(0, \mu, \beta_0(\mu)) > 0.$$

As $\beta_0(\mu)$ is chosen optimally, $(y - \beta_0(\mu))\,\hat{\lambda}(0, \mu, \beta_0(\mu)) > (y - \beta_1(\mu))\,\hat{\lambda}(0, \mu, \beta_0(\mu))$ and so the left-hand side of the above inequality must be less than

$$(y - \beta_1(\mu))\,\hat{\lambda}(1, \mu, \beta_1(\mu)) - c - (y - \beta_1(\mu))\,\hat{\lambda}(0, \mu, \beta_0(\mu)).$$

This simplifies to

$$(y - \beta_1(\mu))\left[\hat{\lambda}(1, \mu, \beta_1(\mu)) - \hat{\lambda}(0, \mu, \beta_0(\mu))\right] - c.$$

As this must be positive, this implies that $\hat{\lambda}(1, \mu, \beta_1(\mu)) > \hat{\lambda}(0, \mu, \beta_0(\mu))$.∎

Intuitively this is because the firm has to cover the cost of being mission oriented which only happens if this induces more effort and hence higher output. To prove our next result, we make the following assumption:

Assumption 2

$$p\left[F\left(p\hat{\beta}_1(1) + \theta\right)\left[y - \hat{\beta}_1(1)\right] - F\left(p\hat{\beta}_0(1) - \varepsilon\right)\left[y - \hat{\beta}_0(1)\right]\right] > c$$

This always holds if c is small enough. We now have the following result comparing profits across the two choices available to the firm:

Proposition 3 *Suppose that Assumption 2 holds, then there exists $\tilde{\mu}$ such that the firm uses pro-social motivation if and only if $\mu \geq \tilde{\mu}$.*

This says that firms will become mission-oriented in order to motivate their workers if it is cheap enough to do so and there are sufficiently many motivated workers. There is a sense in which the workers exercise soft power over the firm's objective and to what would look to an outsider as curtailing profit maximization. However, observing that profits are lower because c is incurred would constitute a misunderstanding of the bigger picture and the fact that it is indeed optimal to keep an existing work force of motivated agents happy.

Now let us consider what happens to the payoffs of each type. For this, observe that

$$\Delta(\mu) = \begin{cases} [F(p\beta_1(\mu) + \theta) - F(p\beta_1(\mu))]\,p\beta_1(\mu) > 0 & \mu \geq \tilde{\mu} \\ [F(p\beta_0(\mu) - \varepsilon) - F(p\beta_0(\mu))]\,p\beta_0(\mu) < 0 & \text{otherwise.} \end{cases} \tag{8}$$

16

This implies that which type has a fitness advantage depends on whether the organization chooses to be mission-oriented. Note that for all $\mu \in [0, 1]$, $\Delta(\mu)$ is therefore strictly positive or negative which rules out the possibility of an interior steady state.

Multiple Steady States   Next, we explore how this affects the cultural dynamics. We have the following result:

Proposition 4 *For all $\mu_0 \geq \tilde{\mu}$, the organization converges to a stable steady state where $\mu = 1$ in the long run. Otherwise, the only stable steady state has $\mu = 0$.*

The evolutionary path is pinned down directly by the organizational choice which itself depends on the initial condition μ_0 and there are now multiple steady states depending on the starting point. If the starting value of μ is high enough, then the organization will choose a mission to suit motivated agents which creates an efficiency advantage and economizes on monetary incentives. This will result in a psychological fitness advantage to motivated agents which means that there number of increases over time until reaching its high steady state value. The converse is true when the organization begins with a low value of μ. This will result in a fall in the proportion of motivated agents until the organization is populated exclusively by selfish agents. The reason for this is that motivated types put in higher effort compared to selfish types when $\sigma = 1$ and the converse is true when $\sigma = 0$. Hence there is a complementarity between μ and the socialization process.

A Role for Motivated Founders?   If $\mu_0 = 0$ is the natural state as in most economic models, then it would seem unlikely to be able to get to a position where $\mu > \mu_0$ since this steady state is always stable. However, we believe that the best explanation for how this happens in practice is by having firms with inspirational mission-oriented founders who are themselves willing to sacrifice profits for wider goals and who are able to socialize an initial group of workers into being motivated and inspire at least a small group of workers in an organization to become motivated. Our model then predicts that this initial condition can result in hysteresis where the firm continues to subscribe to the founder's values after the founder has left the firm.

To see this, suppose that there is a founder with a preference for $\sigma = 1$ i.e. who gets utility $\Theta > c$ from choosing this. Suppose that Θ is large enough such that

$$\hat{\Pi}(0,1) = \max_{\beta \geq 0} \{[y - \beta] pF(p\beta) + \Theta - c\} > \hat{\Pi}(0,0) = \max_{\beta \geq 0} \{[y - \beta] pF(p\beta)\}.$$

This implies that the steady state at $\mu^* = 0$ is no longer stable since $\Delta(\nu) > 0$ as $\nu \to 0$. Hence any small random perturbation which results in a few workers becoming motivated will "infect" the population and to motivation in the organization growing over time. Hence, a process of socialization will gradually lead to μ growing over time according to the process in (7). If the founder stays in charge for ever, this would lead to $\mu = 1$ being the steady state. Thus, we have:

Proposition 5 *Suppose that the founder is motivated, then the unique stable steady state when the founder is in charge is $\mu = 1$.*

However, we do not require that there is a permanently motivated founder in charge. Even after the founder is no directly involved in the firm, then this can shape the future trajectory of an organization long after he/she has left the firm and its taken over by a purely profit maximizing owner as long as at the point of his departure from the organization, $\mu > \tilde{\mu}$. However, if the takeover is too early in a firm's history, then a profit-maximizing leader would revert to $\sigma = 0$ and the firm would converge back to $\mu = 0$.

An example which fits this path is Ben and Jerry's Ice Cream which was established by two motivated founders with certain principles about ethical sourcing of inputs. However, it was then taken over by Unilever. Many at the time doubted whether this would result in the its ethical stance being maintained. However, so far this appears to have been the case. This is consistent with the model as it would be optimal to adhere to the mission by a profit-maximizer once it is entrenched making it credible that the non-profit mission is preserved. Moreover, the apparent non-profit mission will actually generate higher profits than a for-profit mission due to the motivational benefits on employees. Hence, it is indeed optimal for Unilever to maintain the founder's mission even if it does not directly subscribe to it.

Regulation   Another way of effecting change and creating an incentive for some forms of motivation is via regulation. Suppose that regulation forces a

firm to pick a specific mission such as mandating a green technology. The classic view is that such regulation is always efficiency reducing. And in the near term with $\mu = 0$, this is surely the case. As with motivated founders, sustained regulation can lead to change since the steady state where $\mu = 0$ is no longer stable.

Proposition 6 *Suppose that there is a regulation that sets $\sigma = 1$. Then the unique stable steady state while the regulation is in place is $\mu = 1$.*

This suggests an intriguing possibility that having a firm which becomes greener can actually given actually lead workers to value this stance and that this could enhance worker motivation, mitigating the efficiency loss in firms. Giving a fitness advantage to environmentally motivated workers whose payoffs are now higher may eventually bring in more profit in the long-run as workers are willing to work hard for green firms. Moreover, it is self-sustaining in the sense that it becomes optimal for the firm to maintain a green stance even if the regulation us taken away.

This creates an argument for creating regulation strategically to effect cultural change. Of course, this can also take darker forms. For example, countries with strong ideologies such as communist countries like North Korea engage in intense propaganda for national loyalty. This may have an impact on long run productivity reducing the need for bonus pay. But it would have grim consequences for a transition from socialization when there would be a loss of efficiency as firms can no longer draw on this form of motivation.

How Societal Change can Change Business Practice We have assume that everyone comes to the firm as selfish and then gets socialized in the work place. But there is the possibility that the precursor to such cultural change in business is a new workforce which is socialized in to different practices. Hence another way to change μ is for the firm to be hiring a cadre of workers who begin life as motivated when they enter the firm due to their schooling or upbringing being different from existing workers. Having women enter the labor force in larger numbers with new attitudes to work place practices can also be important and create a group of workers, for for example, with workers who prefer to work in an environment free from sexual harassment.

In practice, we know that some existing workers may try to socialize away from these "new" views. But overtime, they could become entrenched

depending the threshold value of $\tilde{\mu}$. This may then require some amount of regulation as discussed in the section above. However, it may be that infecting worker environments with non-sexist workers is sufficient to create a non-sexist firms in the long-run.

Recall that every period, a fraction ρ of the workers will be replaced. And we suppose that additionally, among the new workers, $\hat{\mu}$ will be mission-motivated and remains so after socialization. Then we have

$$\mu_{t+1} = (1 - \rho)\mu_t + \rho\hat{\mu} + \rho(1 - \hat{\mu}) \left[2\mu_t(1 - \mu_t)G(\Delta(\mu_t)) + \mu_t^2 \right]$$

Rearrange the terms, we have

$$
\begin{aligned}
\mu_{t+1} - \mu_t &= \rho \left[-\mu_t + \hat{\mu} + (1 - \hat{\mu}) \left[2\mu_t(1 - \mu_t)G(\Delta(\mu_t)) + \mu_t^2 \right] \right] \\
&= \mu_t(1 - \mu_t) \left[2G(\Delta(\mu_t)) - 1 \right] - \hat{\mu}\mu_t(1 - \mu_t) \left[2G(\Delta(\mu_t)) - 1 \right] + \hat{\mu}(1 - \mu_t) \\
&= (1 - \mu_t) \left[(1 - \hat{\mu}) \mu_t [2G(\Delta(\mu_t)) - 1] + \hat{\mu} \right]
\end{aligned}
$$

Then to find steady state, impose $\mu_{t+1} - \mu_t = 0$, we immediately have one steady state being $\mu = 1$, and the other steady state (if it exists) should satisfy

$$G\left(\Delta\left(\mu\right)\right) = \frac{1}{2} \left[1 - \frac{\hat{\mu}}{(1 - \hat{\mu})\mu} \right] < \frac{1}{2}.$$

This implies that the steady state $\mu \leq \tilde{\mu}$ (if it exists), i.e. firm will not use pro-social motivation. We could also explicitly write down $\Delta(\mu)$ as

$$\Delta(\mu) = \left[F\left(p\beta_0(\mu) - \varepsilon\right) - F\left(p\beta_0(\mu)\right) \right] p\beta_0(\mu)$$

Hence the interior steady state μ (if it exists) is characterized by

$$G\left[\left[F\left(p\beta_0(\mu) - \varepsilon\right) - F\left(p\beta_0(\mu)\right) \right] p\beta_0(\mu) \right] = \frac{1}{2} \left[1 - \frac{\hat{\mu}}{(1 - \hat{\mu})\mu} \right].$$

Notice that as $\mu \to 0$, the left hand side of the above equation remains positive but the right hand side goes to $-\infty$. Hence a sufficient condition that guarantees the existence and stability (and uniqueness) of the interior steady state where firms do not use pro-social motivation would be (i) $\Delta(\mu)$ is monotonic for $\mu \in [0, \tilde{\mu}]$, and (ii) $G(\Delta(\tilde{\mu})) \leq \frac{1}{2} \left[1 - \frac{\hat{\mu}}{(1-\hat{\mu})\tilde{\mu}} \right]$.

Competition   We now consider how productivity shocks influence whether firms become mission-orientated. We model this by considering a change in y. Note that

$$\left(y-\beta_1\left(\mu\right)\right)\hat{\lambda}\left(1,\mu,\beta_1\left(\mu\right)\right)-c-\left(y-\beta_0\left(\mu\right)\right)\hat{\lambda}\left(0,\mu,\beta_0\left(\mu\right)\right)$$

is increasing in y since, as we observed above, output is higher in mission oriented firms. So $\tilde{\mu}$ is decreasing in y. Hence there is a greater chance that firms will become mission oriented in a world where financial returns are higher. This could be because of higher productivity per worker or because there is less competition and hence profitability is higher. Thus, there maybe a trade-off between greater competition and being mission oriented all else equal.

Material versus Psychological Fitness   We have taken a materialistic view of fitness as a driver of motivation. However, our results also hold if we suppose that the process of socialization is based on psychological rather than material fitness. The only difference between these two cases is that utility includes the cost of effort associated in achieving rewards not just the payoffs.

   With $\sigma=1$, then the utility difference between being motivated and selfish is

$$
\begin{aligned}
e_1\left(\mu\right) &= \int_0^{p\hat{\beta}_1(\mu)+\theta}\left[p\hat{\beta}_1\left(\mu\right)+\theta-\psi\right]dF\left(\psi\right)-\int_0^{p\hat{\beta}_1(\mu)}\left[p\hat{\beta}_1\left(\mu\right)-\psi\right]dF\left(\psi\right) \\
&= \theta F\left(\hat{\beta}_1\left(\mu\right)p+\theta\right)+\int_{p\hat{\beta}_1(\mu)}^{p\hat{\beta}_1(\mu)+\theta}\left[p\hat{\beta}_1\left(\mu\right)-\psi\right]dF\left(\psi\right)>0
\end{aligned}
$$

and with $\sigma=0$, it is

$$
\begin{aligned}
e_0\left(\mu\right) &= \int_0^{p\hat{\beta}_0(\mu)-\varepsilon}\left[p\hat{\beta}_0\left(\mu\right)-\varepsilon-\psi\right]dF\left(\psi\right)-\int_0^{p\hat{\beta}_0(\mu)}\left[p\hat{\beta}_0\left(\mu\right)-\psi\right]dF\left(\psi\right) \\
&= -\varepsilon F\left(p\hat{\beta}_0\left(\mu\right)-\varepsilon\right)-\int_{p\hat{\beta}_0(\mu)-\varepsilon}^{p\hat{\beta}_0(\mu)}\left[p\hat{\beta}_0\left(\mu\right)-\psi\right]dF\left(\psi\right)<0
\end{aligned}
$$

And hence

$$\Delta\left(\mu\right)=\left\{\begin{array}{ll} e_1\left(\mu\right) & \text{if } \mu\geq\tilde{\mu} \\ e_0\left(\mu\right) & \text{otherwise.}\end{array}\right.$$

So the core dynamics associated with (7) are essentially the same in this case.

Welfare   We now consider the implications for welfare of having more motivated agents in an organization. The exact welfare criterion to use is not entirely clear. First, there is the welfare of workers and profits of the firm. Second, there are any wider benefits associated with the firm choosing $\sigma = 1$. If for example, the action that appeals to agents is genuinely pro-social and creates a benefit to society of $S > c$, then there can be an overwhelming case for having more workers choose this. Of course, in such cases, it would be natural for $\sigma = 1$ to be imposed by regulation. Hence, studying the decentralized solution applies only when the government chooses not to act or the monitoring costs of a regulatory intervention are very high.

For the moment we set aside wider social benefits and study only welfare within an organization may also be of interest. By construction, we have suppose that there is a utility from being motivated of θ. Hence, this will always increase the welfare of workers. The firm will also only pick $\sigma = 1$ if it is more profitable for it to do so.

To explore welfare, we consider, the total surplus in the firm the sum of profits and worker utility. Recall the expression of welfare when $\mu = 1$ is

$$W_1 = \int_0^{p\hat{\beta}_1(1)+\theta} [py + \theta - \psi]dF(\psi) - c,$$

and the expression can be written as

$$W_0 = \int_0^{p\hat{\beta}_0(0)} [py - \psi]dF(\psi).$$

Next we try to work out an explicit expression for $W_1 - W_0$, and determine under what circumstances this is positive. To explicitly write down $W_1 - W_0$, we discuss two cases, i.e. $\theta > p[\hat{\beta}_0(0) - \hat{\beta}_1(1)]$ or $\theta \leq p[\hat{\beta}_0(0) - \hat{\beta}_1(1)]$. We can show that these two cases lead to the same expression for $W_1 - W_0$.

When $\theta > p[\hat{\beta}_0(0) - \hat{\beta}_1(1)]$, we have

$$
\begin{aligned}
W_1 - W_0 &= \int_0^{p\hat{\beta}_0(0)} \theta dF(\psi) + \int_{p\hat{\beta}_0(0)}^{p\hat{\beta}_1(1)+\theta} [py + \theta - \psi]dF(\psi) - c \\
&= \theta F\left(p\hat{\beta}_0(0)\right) + [py + \theta]\left[F\left(p\hat{\beta}_1(1) + \theta\right) - F\left(p\hat{\beta}_0(0)\right)\right] - \int_{p\hat{\beta}_0(0)}^{p\hat{\beta}_1(1)+\theta} \psi dF(\psi) - c \\
&= [py + \theta]F\left(p\hat{\beta}_1(1) + \theta\right) - py F\left(p\hat{\beta}_0(0)\right) - \int_{p\hat{\beta}_0(0)}^{p\hat{\beta}_1(1)+\theta} \psi dF(\psi) - c.
\end{aligned}
$$

When $\theta \leq p[\hat{\beta}_0(0) - \hat{\beta}_1(1)]$, we have

$$
\begin{aligned}
W_1 - W_0 &= \int_0^{p\hat{\beta}_1(1)+\theta} \theta dF(\psi) - \int_{p\hat{\beta}_1(1)+\theta}^{p\hat{\beta}_0(0)} [py - \psi]dF(\psi) - c \\
&= \theta F\left(p\hat{\beta}_1(1) + \theta\right) - py\left[F\left(p\hat{\beta}_0(0)\right) - F\left(p\hat{\beta}_1(1) + \theta\right)\right] + \int_{p\hat{\beta}_1(1)+\theta}^{p\hat{\beta}_0(0)} \psi dF(\psi) - c \\
&= [py + \theta]F\left(p\hat{\beta}_1(1) + \theta\right) - py F\left(p\hat{\beta}_0(0)\right) - \int_{p\hat{\beta}_0(0)}^{p\hat{\beta}_1(1)+\theta} \psi dF(\psi) - c,
\end{aligned}
$$

which is exactly the same expression as with $\theta > p[\hat{\beta}_0(0) - \hat{\beta}_1(1)]$. Hence $W_1 \geq W_0$ if and only if

$$
[py + \theta]F\left(p\hat{\beta}_1(1) + \theta\right) - py F\left(p\hat{\beta}_0(0)\right) - \int_{p\hat{\beta}_0(0)}^{p\hat{\beta}_1(1)+\theta} \psi dF(\psi) \geq c, \qquad (9)
$$

provided that Assumption 2 holds.

5 Concluding Comments

This paper has put forward a framework for studying the evolution of motivation alongside the reward structures in organizations. It has emphasized how these co-evolve and that the choice of reward structures can either enhance or diminish motivation. We have shown that organizations can harness non-pecuniary motivations even when the goal of the organization is profit-maximization. However, there is natural threshold effects which means that this takes hold only when μ is sufficiently high. Otherwise, there is a move towards standard selfish preferences.

The paper fits into a wider agenda which appreciates that the profit motive has wider consequences for the culture of societies as emphasized, for example, by Sandel (2012) and Titmuss (1970) in an earlier era. But to appreciate their arguments, it is necessary to utilize the idea that motivation is endogenous. Although this remains a debating point in economics, the tools developed, for example, in Alger and Weibull (2013), Dekel et al (2007) and Sethi and Somanathan (2001) open up these possibilities. Putting structure on this also helps to give some discipline to the process of preference change and illustrates the limits on the arguments. It also illustrates the

range of circumstances in which *homo economicus* has a fitness advantage and hence the incentives that society uses become more like those that appear in textbook economic models.

References

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

[2] Akerlof, George, and Janet L. Yellen, [1990], "The Fair Wage-Effort Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105(2), 255-283.

[3] Alger, Ingela and Jorgen Weibull, [2013], "Homo Moralis – Preference Evolution Under Incomplete Information and Assortative Matching," *Econometrica*, 81(6), 2269–2302.

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

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

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

[7] Besley, Timothy and Maitreesh Ghatak, [2016], "Market Incentives and the Evolution of Motivation," unpublished typescript.

[8] Besley, Timothy and Torsten Persson, [2019], "Democratic Values and Institutions", American Economic Review - Insights, 1(1), 59-76.

[9] Besley, Timothy and Torsten Persson, [2022], "Organizational dynamics: culture, design, and performance" Forthcoming, *The Journal of Law, Economics, and Organization.*

[10] Bidner, Chris and Patrick Francois, [2012], "Cultivating Trust Norms, Institutions and the Implications of Scale," *Economic Journal*, 121, 1097-1129.

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

[12] 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.

[13] Bisin, Alberto and Thierry Verdier (forthcoming),"Advances in the Economic Theory of Transmission ". *Journal of Economic Literature*.

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

[15] Bowles, Samuel and Herbert Gintis, [2011], *A Cooperative Species: Human Reciprocity and Its Evolution*, Princeton: Princeton University Press.

[16] Carpenter, Jeffrey and Erick Gong (2016),"Motivating Agents: How Much Does the Mission Matter? ". *Journal of Labor Economics*, 34(1), 211-236.

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

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

[19] Deci Edward L., R. Koestner R and Richard M. Ryan, [1999], "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," Psychological Bulletin, 125(6), 627-68.

[20] Dekel, Eddie, Jeffrey Ely and O. Yilankaya, [2007], "Evolution of Preferences," *Review of Economic Studies*, 74, 685–704.

[21] Deserranno, Erika [2017], "Financial incentives as signals: Experimental evidence from the recruitment of village promoters in Uganda," working paper.

[22] Dessein, Wouter and Andrea Prat (2022),"Organizational Capital, Corporate Leadership, and Firm Dynamics ". *Journal of Political Economy*, 130(6), 1477-1536.

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

[24] Fehr, Ernst and Armin Falk, [2002], "Psychological foundations of incentives," *European Economic Review*, 46, 687 – 724.

[25] Fehr, Ernst and Klaus M. Schmidt, [1999], "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114(3), 817-868.

[26] Francois, Patrick and Jan Zabojnik, [2005], "Trust, Social Capital and Economic Development", *Journal of the European Economic Association*, 3(1), 51-94.

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

[28] 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.

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

[30] 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.

[31] Hiller, Victor and Thierry Verdier (2014),"Corporate Culture and Identity Investment in an Industry Equilibrium ". *Journal of Economic Behavior & Organization*, 103, 93-112.

[32] Kamenica, Emir, [2012], "Behavioral Economics and Psychology of Incentives," *Annual Review of Economics*, 4, 427-452.

[33] Kandel, Eugene and Edward Lazear, [1992], "Peer Pressure and Partnerships," *The Journal of Political Economy*, 100(4), 801-817.

[34] Lazear, Edward, [1991], "Labor Economics and the Psychology of Organizations," *The Journal of Economic Perspectives*, 5(2), 89-110.

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

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

[37] 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.

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

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

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

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

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

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

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

Appendix

Proof of Lemma 1 Recall the first order condition yields

$$(y - \beta)\lambda' - \lambda = 0.$$

We can check that the second order condition holds by plugging in the first order condition:

$$\frac{\partial^2 \hat{\Pi}}{\partial \beta^2} = (y - \beta)\lambda'' - 2\lambda' = \frac{1}{\lambda'}[\lambda''\lambda - 2\lambda'^2].$$

We know this is negative from log concavity of λ in β.

To show the comparative static results, we can obtain from the first order condition that

$$\frac{\partial \beta(\mu, v)}{\partial \mu} = -\frac{(y - \beta)\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial\mu}}{\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta^2} - \frac{\partial \log \lambda(v,\mu,\beta)}{\partial\beta}},$$

and

$$\frac{\partial \beta(\mu, v)}{\partial v} = -\frac{(y - \beta)\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial v}}{\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta^2} - \frac{\partial \log \lambda(v,\mu,\beta)}{\partial\beta}}.$$

From log concavity of λ, we know $\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta^2} < 0$. What remains is to determine the signs of $\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial\mu}$ and $\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial v}$. We have

$$
\begin{aligned}
\frac{\partial^2 \log \lambda(v, \mu, \beta)}{\partial\beta\partial\mu} &= \frac{p}{\lambda^2}[[F'(\beta p + v) - F'(\beta p)][\mu F(\beta p + v) + (1 - \mu)F(\beta p)] \\
&\quad - [F(\beta p + v) - F(\beta p)][\mu F'(\beta p + v) + (1 - \mu)F'(\beta p)]] \\
&= \frac{p}{\lambda^2}\left[F'(\beta p + v)F(p\beta) - F'(\beta p)F(\beta p + v)\right].
\end{aligned}
$$

Hence $\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial\mu} < 0$ if and only if

$$\frac{F'(\beta p + v)}{F(\beta p + v)} < \frac{F'(\beta p)}{F(p\beta)}.$$

From the log concavity of $F(.)$, we know $\frac{F'(\beta p+v)}{F(\beta p+v)}$ is decreasing in v. Hence the above inequality holds if and only if $v > 0$. Therefore, when $v > 0$,

29

$\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial\mu} < 0$, and when $v < 0$, $\frac{\partial^2 \log \lambda(v,\mu,\beta)}{\partial\beta\partial\mu} > 0$. This implies that when $v > 0$, $\beta_1(\mu, v)$ is decreasing in μ, and when $v < 0$, $\beta_0(\mu, v)$ is increasing in μ.

To show $\beta_1(\mu, v)$ is decreasing in v (as $\sigma = 1$, in this case $v = \theta > 0$), again we examine

$$
\begin{aligned}
\frac{\partial^2 \log \lambda(v, \mu, \beta)}{\partial\beta\partial v} &= \frac{p}{\lambda^2}[[\mu F''(\beta p + v)][\mu F(\beta p + v) + (1 - \mu)F(\beta p)] \\
&\quad - [\mu F'(\beta p + v)][\mu F'(\beta p + v) + (1 - \mu)F'(\beta p)]] \\
&\leq \frac{p\mu}{\lambda^2}[\frac{F'(\beta p + v)^2}{F(\beta p + v)}[\mu F(\beta p + v) + (1 - \mu)F(\beta p)] \\
&\quad - [F'(\beta p + v)][\mu F'(\beta p + v) + (1 - \mu)F'(\beta p)]] \\
&= \frac{p\mu}{\lambda^2}F'(\beta p + v)(1 - \mu)F(\beta p)\left[\frac{F'(\beta p + v)}{F(\beta p + v)} - \frac{F'(\beta p)}{F(\beta p)}\right] < 0,
\end{aligned}
$$

given log-concavity of $F(.)$, since $v = \theta > 0$. Then it follows that $\beta_1(\mu, v)$ is decreasing in v.

To show that $\beta_0(\mu, v) > 0$ for all $\mu \in [0, 1]$, note first that $\beta_0(\mu, v) \geq \beta_0(0, v) = \beta_0(0)$. From the first order condition, we know that $\beta_0(0)$ is characterized by

$$
\frac{1}{y - \beta} = p\frac{F'(\beta p)}{F(\beta p)}
$$

As $\beta \to 0$, the left hand side goes to $\frac{1}{y}$ and the right hand side goes to infinity given that $F(0) = 0$. By Assumption 1 $F'(0)$ is bounded. As $\beta \to y$, the left hand side goes to infinity, and the right hand side goes to a constant, i.e. $p\frac{F'(py)}{F(py)} > 0$. Hence there must exist a unique $\beta \in (0, y)$ that solves the above equation.

To show that $\beta_1(\mu, v) > 0$ for all $\mu \in [0, 1]$, again notice that $\beta_1(\mu, v) \geq \beta_1(1, \theta)$. From the first order condition, we know that $\beta_1(1, \theta)$ is characterized by

$$
\frac{1}{y - \beta} = p\frac{F'(\beta p + \theta)}{F(\beta p + \theta)}
$$

Similar to the argument above, as β goes from 0 to y, the left hand side of the above equation increases from $\frac{1}{y}$ to infinity, whereas the right hand side decreases from $p\frac{F'(\theta)}{F(\theta)}$ to $p\frac{F'(py+\theta)}{F(py+\theta)}$. Hence the above equation yields a positive

solution to $\beta_1(1, \theta)$ if and only if

$$\frac{1}{py} < \frac{F'(\theta)}{F(\theta)}.$$

Denote $h(a) = \frac{F'(a)}{F(a)}$ and from the log-concavity of $F(.)$, the above inequality is equivalent to

$$\theta < h^{-1}\left(\frac{1}{py}\right).$$

Proof of Proposition 3   If $\sigma = 1$

$$\hat{\Pi}(\mu, 1) = \max_{\beta \geq 0} \{[y - \beta] \, p \, [\mu F(p\beta + \theta) + (1 - \mu) F(p\beta)] - c\}$$

and with $\sigma = 0$, it is

$$\hat{\Pi}(\mu, 0) = \max_{\beta \geq 0} \{(y - \beta) \, p \, [\mu F(p\beta - \varepsilon) + (1 - \mu) F(p\beta)]\}.$$

Note that $\hat{\Pi}(0, 1) = \max_{\beta \geq 0} \{[y - \beta] \, pF(p\beta) - c\}$, and note that $\hat{\Pi}(0, 1) = \hat{\Pi}(0, 0)$ when $c = 0$. Thus, $\hat{\Pi}(0, 1) < \hat{\Pi}(0, 0)$ for all $c > 0$.
    If $\mu = 1$:

$$\begin{aligned}
\hat{\Pi}(1, 1) &= \left[F\left(p\hat{\beta}_1(1) + \theta\right)\right] p \left[y - \hat{\beta}_1(1)\right] - c \\
&\geq F\left(p\hat{\beta}_0(1) + \theta\right) p \left[y - \hat{\beta}_0(1)\right] - c \\
&> F\left(p\hat{\beta}_0(1) - \varepsilon\right) p \left[y - \hat{\beta}_0(1)\right] - c = \hat{\Pi}(1, 0) - c
\end{aligned}$$

for $\theta > 0$ by the fact that $F(.)$ is increasing where the first inequality holds since $\hat{\beta}(1)$ is the profit maximizing bonus. Hence there exists a range of $c \in [0, \bar{c}]$ where $\bar{c} > 0$ such that:

$$\hat{\Pi}(1, 1) > \hat{\Pi}(1, 0).$$

    Finally note that for all $\mu \in [0, 1]$,

$$\frac{d\left[\hat{\Pi}(\mu, 1) - \hat{\Pi}(\mu, 0)\right]}{d\mu} = \begin{aligned} & p\left[F\left(p\hat{\beta}_1(\mu) + \theta\right) - F\left(p\hat{\beta}_1(\mu)\right)\right]\left[y - \hat{\beta}_1(\mu)\right] \\ & -p\left[F\left(p\hat{\beta}_0(\mu) - \varepsilon\right) - F\left(p\hat{\beta}_0(\mu)\right)\right]\left[y - \hat{\beta}_0(\mu)\right] > 0 \end{aligned}$$

given that $\hat{\beta}_1(\mu) < \hat{\beta}_0(\mu)$. Given the values of $\hat{\Pi}(\mu, 1)$ and $\hat{\Pi}(\mu, 0)$ at $\mu = 0$ and $\mu = 1$, as well as the continuity and monotonicity of $\hat{\Pi}(\mu, 1) - \hat{\Pi}(\mu, 0)$, this establishes the fact that there exists $\tilde{\mu}$ such that the firm uses pro-social motivation if and only if $\mu \geq \tilde{\mu}$.∎

Proof of Proposition 4 Using (7) and (8), $\mu \geq \tilde{\mu}$, then $\Delta(\mu) > 0$ and $\mu_{t+1} - \mu_t > 0$. However if $\mu < \tilde{\mu}$, then $\Delta(\mu) < 0$ and $\mu_{t+1} < \mu_t$. This implies that $\mu = 1$ and $\mu = 0$ are both stable steady states. Given the choice of σ by the firm, $\Delta(\mu) = 0$ cannot occur: when the firm switches from $\sigma = 1$ to $\sigma = 0$, $\Delta(\mu)$ turns strictly negative from strictly positive. Therefore, there is no other interior steady state.∎