The Evolution of Motivation*

Timothy Besley
LSE and CIFAR

Maitreesh Ghatak
LSE

September 13, 2017

Abstract

This paper studies the idea that reward structures in organizations co-evolve with the motivation of workers. Firms employ workers and choose an incentive scheme. New workers who join an organization are socialized by those who work there already. We study three examples of non-pecuniary motivation: pure intrinsic motivation, mission-based motivation and motivation based on reciprocity. In all three cases there is a complementarity between limiting the use of pecuniary rewards, specifically bonus pay, and creating an incentive to become a motivated agent. We characterize the conditions under which non-pecuniary motivation grows or diminishes over time.

*We are grateful to Jay Lee for extremely useful insights and comments on an earlier draft. Email address: t.besley@lse.ac.uk and m.ghatak@lse.ac.uk.

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 work place 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?

We explore three main strands of the motivation literature that we consider all of which have been discussed in psychology and experiments. These are intrinsic motivation, mission motivation and motivation by fairness in rewards. We give conditions under which an organization can sustain such forms of non-pecuniary motivation given that motivation levels are endogenous. One of the key issues is how far market economies undermines motivation, i.e., when organizations are run purely on profit maximizing lines. We emphasize that it is the way that rewards are organized and not their level which is key to the evolutionary process as this determines the comparative fitness of the motivated types in the population.

One of our key findings is, the long-term survival of non-pecuniary motivation in the population depends crucially on whether more motivated agents receive a higher payoff that gives them a fitness advantage in the cultural evolution dynamics. This in turn depends on the exact micro-foundation of the source of the non-pecuniary motivation. If all that being motivated means is not valuing money enough relative to selfish agents, which is how we formulate intrinsic motivation, then it will not survive in the long-run. In contrast, if motivation derives from having pro-social preferences then under appro-

priate parameter conditions, it can survive in the long-run. Interestingly, if motivation derives from workers valuing reciprocity from their co-workers in terms of putting in a fair share of effort, and are willing to inflict non-pecuniary punishments on their selfish peers if they shirk, then motivation also can survive in the long-run. This is because being selfish has a cost that lowers payoffs, which translates into a fitness disadvantage for that type.

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 non-selfish 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 three applications of the ideas and in section five we discuss some extensions. Section six offers concluding remarks.

2 Related Literature

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).

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 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.

Economists have studied the origins and implications of intrinsic motivation in economic settings.[1] 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" instead of requiring that 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 labor 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,

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

use of incentive pay and productivity could be negatively correlated.[2] 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 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, 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).

Mission motivation This follows the approach suggested in Besley and Ghatak (2005) where workers are typically *motivated agents,* i.e. agents who pursue goals because they perceive intrinsic benefits from doing so. 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 be-

[2] See also Deserranno (2017) for a model along these lines.

cause individuals care not just about the levels of these collective goods, but their personal involvement in their production (i.e., a "warm glow").

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).

Reciprocity There is a large literature which argues that a sense of fairness underpinned by some form of reciprocity is powerful in human societies. Biologists such as Trivers (1971) developed the logic reciprocal motives in an evolutionary framework. Nowak (2006) discusses the role of reciprocal motives in the evolution of cooperation. Within economics Bowles and Gintis (2011) pull together a range of research from economics, anthropology and biology on the power of reciprocity in sustaining cooperation.

The idea that a preference for fairness is important in affecting incentives has been developed in Akerlof and Yellen (1991), Fehr and Falk (2002) and Fehr and Schmidt (1999). These ideas have also been developed in game theoretic settings. For example, Rabin (1993) develops a framework in which where the motivations of people to sacrifice their own well-being to help those who are being kind and to punish those who are being unkind drive fairness equilibria in strategic settings. A core idea in such models is that individuals are guided by norms of fairness and are willing to punish those who violate the norms even when it is costly to themselves. Sobel (2005) reviews some of the main ideas involved and how the idea of intrinsic reciprocity, i.e. as part of preferences can be modeled. This is motivated by the voluminous experimental literature which finds strong evidence of reciprocity at work in lab settings. Levine (1998) and Bolton and Ockenfels (2000) develop theoretical models with preference formulations to capture

reciprocity and equity to explain a range of lab experiments using.

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).[3] 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.[4] Bidner and Francois (2012) develop a general equilibrium where norm-driven behavior evolves endogenously.

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.

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

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

In that paper, competition among firms for workers whose types are subject to asymmetric information, and how the evolution of motivation affects overall productivity are the key distinguishing features relative to the present exercise. Besley and Persson (2017) study the interaction between the organization of production in terms of degree of centralization, and organizational culture and how they coevolve in an overlapping-generations model where junior managers are socialized by senior managers.

3 Core framework

In the core model, an organization operates in isolation of others. We can think of this as a labor market where an organization has 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 subsistence self-employment activity that yields an expected return $z \geq 0$. It is the same outside option that the firm takes into account when recruiting new agents. For simplicity, we set $z = 0$.

An organization comprises a continuum of agents indexed $i \in [0, 1]$ each of which is one of two types $\varphi \in \{m, s\}$ 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 $\delta \in \{0, 1\}$. Let individual output be $x(\delta)$ where $x(0) = 0$ for all $i \in [0, 1]$ and

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

Therefore, contingent on choosing $\delta = 1$, expected output is α.[5]

Let

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

[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 insights of the model.

Expected *total* output is $X(\lambda) = \alpha\lambda$ where λ is the proportion of agents in the organization who set $\delta = 1$. The firm earns a revenue of y per unit of output.

Preferences Preferences are linear in consumption c and depend also on the net cost/benefit of effort δ:

$$U(c, \delta) \equiv c + v(\delta; \varphi, \psi, \sigma, \mu). \tag{1}$$

The parameter ψ captures disutility of effort and σ is a parameter that indicates whether the firm uses incentive pay or not. The function $v(\delta; \varphi, \psi, \sigma, \mu)$ also depends on the type of the agent (φ) and the proportion of of motivated workers in the workforce. Our non-standard formulation allows the function $v(\cdot)$ to depend on the agent's type φ (selfish or motivated) and, in general, we expect $v(1; m, \psi, \sigma, \mu) - v(0; m, \psi, \sigma, \mu) \geq v(1; s, \psi, \sigma, \mu) - v(0; s, \psi, \sigma, \mu)$, i.e. an agent who is motivated is more inclined to put in effort.

We assume $v(\delta; \varphi, \psi, \sigma, \mu)$ is non-increasing in ψ, reflecting the standard properties associated with disutility of effort. In addition, we assume $v(1; \varphi, \psi, \sigma, \mu) - v(0; \varphi, \psi, \sigma, \mu)$ is decreasing in $\psi \in \left[0, \bar{\psi}\right]$ where $\bar{\psi} > y$. The latter implies that some agents would not put in effort even if they were made full residual claimants on the additional output that they can produce by putting in effort. The parameter ψ is assumed to have the distribution $F(\psi)$.

The payoff $v(\delta; \varphi, \psi, \sigma, \mu)$ depends on on δ i.e. whether or not the agent is putting in effort. The standard economic assumption would be where $v(\delta; \varphi, \psi, \sigma, \mu)$ is decreasing in δ. And its simplest form in the current model would be simply $v(\delta; \varphi, \psi, \sigma, \mu) = -\psi\delta$. This will be the benchmark *homo economicus* case below.

This formulation of $v(\delta; \varphi, \psi, \sigma, \mu)$ is quite general and allows for a range of possible non-pecuniary motivations. It could, for example, include a concern for the mission of the organization as hypothesized in Besley and Ghatak (2005). Another possibility is a direct concern for procedural fairness for wage setting in the firm as in Akerlof and Yellen (1990). In both cases, this would be captured by the dependence of $v(\delta; \varphi, \psi, \sigma, \mu)$ on σ, the policy decisions made by the firm. It can also depend on μ which should be relevant in a case where having more motivated agents in the firm results in greater pressure on selfish workers if they are susceptible to peer pressure as in the model of Kandel and Lazear (1992).

Remuneration Other than a flat pay component, since projects are iid and there is no direct interaction between the effort decisions of workers (unlike in moral hazard in teams), it is sufficient to concentrate on two outcome measures, aggregate output and individual output. Since this is a large continuum economy, aggregate output $X(\lambda)$ is a constant by the law of large numbers. Since x can take only two values, 0 and 1, and aggregate output is a constant, we can restrict attention to a pay policy of the following form:

$$r(x, X, \sigma) = \omega + \beta x.$$

This formulation allows for a fixed wage, ω, and bonus pay where $\beta > 0$. We will model remuneration rules (σ) as restrictions on $\{\omega, \beta\}$. We represent this idea by supposing that $\{\omega, \beta\} \in C(\sigma)$ for $\sigma \in \{A, B\}$ where $C(\sigma)$ represents the set of contracts that satisfy the relevant remuneration rules. Thus, if some firms are constrained not to offer individual bonus pay, we have a constraint where $\beta(\sigma) = 0$. In what follows, we will suppose that the outside option is normalized to zero but that there is a limited liability constraint such that β and ω are both required to be non-negative.

Optimal Effort Optimal effort solves

$$\hat{\delta}(\varphi, \sigma, \mu, \psi, \beta) = \arg\max_{\delta \in \{0,1\}} \{\beta\alpha\delta + v(\delta; \varphi, \psi, \sigma, \mu)\}.$$

Define $\hat{\psi}(\varphi, \sigma, \mu, \beta) \in [0, \bar{\psi}]$ such that $\hat{\delta}(\varphi, \sigma, \mu, \psi, \beta) = 1$ if and only if $\psi \leq \hat{\psi}(\varphi, \sigma, \mu, \beta)$. We use here the fact that $v(\delta; \varphi, \psi, \sigma, \mu)$ is decreasing in ψ.[6] This encompasses the case where $\hat{\psi}(s, \sigma, \mu, \beta) = 0$ in which case no agents of type s put in effort and $\hat{\psi}(s, \sigma, \mu, \beta) = \bar{\psi}$ in which case, they all do. There is also an "interior" case

$$\beta\alpha + v\left(1; \varphi, \hat{\psi}(\varphi, \sigma, \mu, \beta), \sigma, \mu\right) = v\left(0; \varphi, \hat{\psi}(\varphi, \sigma, \mu, \beta), \sigma, \mu\right)$$

for $\hat{\psi}(\varphi, \sigma, \mu, \beta) \in [0, \bar{\psi}]$. Total effort is:

$$\hat{\lambda}(\sigma, \mu, \beta) = \mu F\left(\hat{\psi}(m, \sigma, \mu, \beta)\right) + (1 - \mu) F\left(\hat{\psi}(s, \sigma, \mu, \beta)\right).$$

[6] Note also that $\hat{\psi}(m, \sigma, \mu, \beta) \geq \hat{\psi}(s, \sigma, \mu, \beta)$ in general, if $v(1; m, \psi, \sigma, \mu) - v(0; m, \psi, \sigma, \mu) \geq v(1; s, \psi, \sigma, \mu) - v(0; s, \psi, \sigma, \mu)$.

Organizational Objectives and Design Each worker produces an output of 1 with probability α if the worker chooses $\delta = 1$. The value of output per unit is y. Therefore, the gross expected revenue to the firm from a given worker is $\alpha\delta y$. As far as bonus pay based on individual output (βx) is concerned, if output is 1 the worker gets paid β (when $\delta = 1$ is chosen) and zero otherwise Therefore, the expected payment to the worker on this account is $\alpha\beta\delta$. Total expected pay to the worker is $\omega + \alpha\beta\delta$. Per worker, the firm's net payoff is, therefore:

$$\pi = (y - \beta)\,\delta\alpha - \omega.$$

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

$$\Pi(X, b, \beta) = [y - \beta]\,X(\lambda) - \omega.$$

Let $\left\{\hat{\omega}_\sigma(\mu), \hat{\beta}_\sigma(\mu)\right\}$ denote the set of optimal remuneration contracts and let

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

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 $\omega = 0$. Hence, the only optimization decision for firms is over β.

We make the following regularity assumption which covers all of the applications below:[7]

Assumption 1: *For all $\nu > 0$, $\gamma \leq 1$, and $\mu \in [0, 1]$ we assume that : (i) $\log(\mu F(\nu + \gamma\alpha\beta) + (1 - \mu) F(\alpha\beta))$ is concave; and (ii) y is large enough so that the function $(y - \beta)[\mu F(\nu + \alpha\gamma\beta) + (1 - \mu) F(\alpha\beta)]$ is strictly increasing in β at $\beta = 0$.*

Note that log-concavity of $\mu F(\nu + \gamma\alpha\beta) + (1 - \mu) F(\alpha\beta)$ is weaker than requiring that $F(\cdot)$ be concave. The following result, whose result is in the Appendix, is used repeatedly below:

Lemma 1 *Suppose that Assumption 1 holds. Then the maximization problem $\max_{\beta\geq 0}[(y - \beta)\,\alpha\{\mu F(\nu + \gamma\alpha\beta) + (1 - \mu) F(\alpha\beta)\}]$ is well-behaved.*

[7] The assumption holds, for example, if the function $F(\cdot)$ is an exponential or Pareto distribution.

As we show in the appendix, Assumption 1 implies that the profit function is quasi-concave and has an optimum where a bonus is used, i.e. $\beta > 0$. Two of our applications will be to contrast individual bonus rewards with flat wages so it will be useful to have a case where standard profit maximizing assumptions guarantee that bonuses are used.

The organization design maximizes the leader's payoff, i.e.,

$$\hat{\sigma}(\mu) = \max_{\sigma \in \{A,B\}} \hat{\Pi}(\mu, \sigma).$$

The interesting possibility raised by the applications below is that either restricting or enhancing the type of remuneration contract that a firm uses can be optimal in order to maximize profits. One feature of the formulation which the reader should bear in mind is that we are allowing σ to enter the function $v(\delta; \varphi, \psi, \sigma, \mu)$ e.g. to represent the worker's views about the merits different contracts so there can be an effect operating on the choice motivation of the type m agents. The applications will be precisely about this and its consequences for the evolution of motivation.

Socialization At the optimal effort level, expected payoffs to each type are summarized in:

$$\begin{aligned} &V(\varphi, \sigma, \mu) = \hat{\omega}_\sigma(\mu) \\ &+E\left\{\left[\hat{\beta}_\sigma(\mu)\,\alpha\hat{\delta}\left(\varphi, \sigma, \mu, \psi, \hat{\beta}_\sigma(\mu)\right) + v\left(\varphi, \sigma, \mu, \psi, \hat{\delta}\left(\varphi, \sigma, \mu, \psi, \hat{\beta}_\sigma(\mu)\right)\right)\right]\right\} \end{aligned}$$

where $E\{\cdot\}$ is the expectations operator taken with respect to ψ. The socialization process depends on "psychological fitness" of the motivated type which is defined as

$$\Delta(\mu) = V(m, \hat{\sigma}(\mu), \mu) - V(s, \hat{\sigma}(\mu), \mu).$$

This function is critical to the evolution of motivation and we derive it for each application that we develop below.[8]

It is reasonable to suppose that the evolutionary are "Darwinian" in the sense that:

$$\mu_{t+1} - \mu_t = Q(\mu_t, \Delta(\mu_t)) \tag{2}$$

[8] An expression for $\Delta(\mu)$ is given in the proofs of Propositions 2, 4 and 6 below.

where $Q(\mu, \Delta)$ is increasing in Δ so that having a larger fitness advantage increases the proportion of motivated types in the population.[9] We will work with a specific formulation which delivers such dynamics.

Suppose that there is turnover in the organization each period with a fraction ρ of the workers who are replaced each period. In deriving the law of motion for μ_t, we make the further assumption that there is a mutation rate u at which motivated agents become selfish each period and remain so in perpetuity. The timing is such that the fraction of motivated agents at the beginning of period t is $(1-u)\mu_t$. We introduce this feature so that an organization where there is no strict incentive to be motivated gradually converges towards having a population of selfish types. We will be focusing on the case where the mutation rate u is small.

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 $\mu_t(1-u)$, we assume that he may become motivated depending on the relative psychological fitness of motivated and selfish types. Any randomly selected agent becomes motivated through mentoring if:

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

where η is a mean-zero, *symmetrically* distributed idiosyncratic shock with continuous distribution function $G(\cdot)$. 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 $G(\Delta(\mu_t))$.[10]

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

[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.

[10] To illustrate, consider the example of a logistic distribution. The probability of a randomly selected new agent to become motivated through mentoring is, then:

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

Observe that $G(\Delta(\mu_t)) > \frac{1}{2}$ for $\Delta(\mu_t) > 0$, $G(0) = \frac{1}{2}$, and $G'(\Delta(\mu_t)) > 0$.

fraction of such types in the organization, a kind of social learning postulated in much of the cultural-evolution literature. Assuming a linear relation, the probability of indirect socialization becomes $(1 - G(\Delta(\mu_t)))\,\mu_t(1-u)$ where $\mu_t(1-u)$ 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 becomes motivated is:

$$G(\Delta(\mu_t)) + \{1 - G(\Delta(\mu_t))\}\mu_t(1-u). \tag{3}$$

If a new worker is matched with and mentored by a selfish worker, which happens with probability $1-\mu_t(1-u)$, she is never directly socialized into being a motivated type. On the other hand she is 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.

The fraction $G(\Delta(\mu_t))$ of new recruits who do not become selfish in this way, can indirectly become motivated (as above) depending on the aggregate fraction of motivated agents $(\mu_t(1-u))$ in the organization. The resulting probability of becoming motivated is therefore:

$$G(\Delta(\mu_t))\,\mu_t(1-u). \tag{4}$$

Multiplying (3) by μ_t, (4) by $1-\mu_t$, adding, and simplifying the resulting expressions yields the following the equation of motion for the share of motivated types:

$$\mu_{t+1} - \mu_t = \mu_t \Gamma(\mu_t) \tag{5}$$

where

$$\Gamma(\mu) = \rho(1 - \mu(1-u))(1-u)[2G(\Delta(\mu)) - 1] - u \tag{6}$$

From this it is clear that studying the evolutionary dynamics of motivation requires studying the sign of $\Delta(\mu_t)$. We now make the following useful regularity assumption:

Assumption 2: *The function* $\Gamma(\mu)$ *is strictly concave for all* $\mu \in [0,1]$.

This assumption is used to rule out multiple interior steady states in the dynamics explored below.

Remark: Assumption 2 is not particularly restrictive. A sufficient condition for this to hold is: $\Delta(\mu)$ is concave (which is the case with all the applications analyzed below), u is small, and the density function $g(.)$ corresponding to $G(.)$ satisfies the condition $\frac{g'(\cdot)}{g}\Delta(\mu) \leq 1$ which holds for standard symmetric distributions such as the logistic and the continuous uniform.[11]

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. A fraction $u\mu_t$ become selfish.
3. The leader chooses the organizational form $\sigma \in \{A, B\}$
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 condition to hold simultaneously:

$$\text{(i) } \sigma^* = \hat{\sigma}(\mu^*), \text{ (ii) } X^* = X\left(\hat{\lambda}(\sigma^*, \mu^*)\right) \text{ and } \mu^*\Gamma(\mu^*) = 0.$$

[11] It is straightforward to check that the second derivative of $\Gamma(\mu)$ is

$$-2g(\Delta(\mu))\Delta'(\mu)\rho(1-u)\left[2(1-u) - \{1-\mu(1-u)\}\left\{\frac{g'(\Delta(\mu))}{g(\Delta(\mu))}\Delta'(\mu) + \frac{\Delta''(\mu)}{\Delta'(\mu)}\right\}\right.$$

A sufficient condition for this to be negative is that u is small, and that

$$\frac{g'(\Delta(\mu))}{g(\Delta(\mu))}\Delta'(\mu) + \frac{\Delta''(\mu)}{\Delta'(\mu)}$$

is either negative or, if positive, less than 1. If $\Delta(\mu)$ is concave, then $\Delta'(\mu) < \Delta(\mu)$ the condition therefore holds as long as $\frac{g'(\cdot)}{g}\Delta(\mu) \leq 1$.

It is immediate from this that there is always a trivial steady state where $\mu = 0$. However, there can also be steady states where $\mu^* > 0$.

We are interested in how an organization converges to a particular steady state and how the motivation matches that steady state behavior. The following result is useful in the applications that follow:

Lemma 2: *Suppose that Assumption 2 holds, u is small enough, there exists $\hat{\mu}$ such that $\Delta(\hat{\mu}) > 0$, then for all $\mu \geq \hat{\mu}$, there exists $\bar{\mu} < 1$ such that $\lim_{t\to\infty} \mu_t \geq \bar{\mu}$.*

This says that when μ is such that the motivated type enjoys a positive level of fitness and the mutation rate is small then μ converges towards a threshold value of $\bar{\mu}$. Since $u > 0$, this threshold is below one, not agents will be motivated in the long-run. However, it will be close to one for small u.

4 Applications

We now develop three applications to areas where there has been extensive discussion of non-standard motivation. They are all specific variants of the model laid out above. In all three cases, choosing $\sigma = A$ will be a policy which expressly tries to take advantage of some kind of non-standard motivation. In the first, not paying an individual bonus brings forth intrinsic motivation. In the second, the firm chooses a mission for the organization to elicit motivation. And in the third application, the firm will choose a system of fair rewards as the basis of eliciting reciprocal motivation among workers. In all three cases, we contrast this with a standard reward scheme, $\sigma = B$, where the firm pays a standard individualized bonus to elicit effort.

In each case, we characterize the conditions under which a profit-maximizing firm chooses the reward scheme and then look at its implications for the evolution of motivation. The choice of three applications allows us to highlight three possibilities. In the case of pure intrinsic motivation, the prospects for a positive evolution of motivation are not good unless it is characterized by a pure "love of effort". In the case of mission motivation, the fact that workers value the mission of the firm is able to give a boost to the evolutionary prospects for mission motivation provided that rewarding workers in this way is not too costly. In the case of reciprocal motivation, which uses peer punishments, the reasoning is more subtle. Even if the firm wishes to harness

such peer punishments, there is no guarantee that fairness motivation will grow over time since such punishments are costly to negative reciprocators. Notwithstanding, there are conditions under motivation for fairness evolves and affects the wage structure chosen by a profit-maximizing firms.

4.1 Intrinsic motivation

In the most basic intrinsic motivation model there are agents who work for the sake of completing the task because they enjoy the process. For example, psychologists such as Ryan and Deci (2000) define intrinsic motivation as "..doing of an activity for its inherent satisfactions rather than for some separable consequence" (p. 56). We can model agents of type m as intrinsically motivated in this way. In particular, we can stipulate that they have no disutility from putting in effort, while selfish agents do.

In terms of (1) above, suppose, $v(1; m, \psi, A, \mu) = v(0; m, \psi, A, \mu) = 0$ for all $\mu \in [0, 1]$.[12] We also assume, following the classic view of such motivated agents that they are less responsive to incentives than selfish agents. Specifically, we suppose that if they are given a bonus, β, as a reward then they value the bonus as $\gamma\beta$ where $\gamma \leq 1$.[13] In the extreme case $\gamma = 0$, then there is complete motivation crowding when motivated agents are asked to complete a task. For the type s agents, $v(\delta; s, \psi, A, \mu) = -\delta\psi$.

Whether or not this is the correct way of capturing pure intrinsic motivation in our setting is debatable. Drawing on experimental evidence, 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. So we could think of $\sigma = A$ as a situation in which the firm delegates autonomy over task choice in situation where a range of complex tasks are required to produce an output for the firm and only the intrinsically motivated are motivated to find a way of delivering when no reward is specified. Whatever, the micro-foundation, the key to this benchmark model of intrinsic motivation and the results that follow is that there is no positive utility from being intrinsically motivated, just the

[12] Equivalently, we could suppose that they get utility from completing the task which exactly offsets their cost of effort.

[13] This is consistent with Deci et al (1999) which provides a meta study of 128 studies and argument that these support the idea that external rewards crowd out intrinsic motivation.

absence of disutility.

The organization chooses between two rules for remuneration where A pays no performance reward and B is a standard individual bonus pay arrangement which applies through the firm. In the first, the absence of pay-for-performance is an organizational rule to which the firm adheres. Hence $\beta_A(\mu) = 0$ while $\beta_B(\mu)$ is freely chosen.

Let $\hat{\beta}_\sigma(\mu)$ be the optimal level of bonus and let

$$\hat{\Pi}(\mu, \sigma) = \max_{\beta \in C(\sigma)} \left\{ \left[y - \hat{\beta}(\sigma) \right] X \left(\hat{\lambda}(\sigma, \mu, \beta) \right) \right\}$$

be the maximized value of profits in each case.

Recognizing that payments to workers are proportional to total output, profit with $\sigma = A$ is

$$\hat{\Pi}(\mu, A) = y\mu\alpha.$$

and for $\sigma = B$ it is:

$$\hat{\Pi}(\mu, B) = \max_{\beta \geq 0} \left[(y - \beta) \alpha \left\{ \mu F(\gamma\alpha\beta) + (1 - \mu) F(\alpha\beta) \right\} \right].$$

We now have the following result for the choice between the two kinds of remuneration policies:

Proposition 1 *There exists $\tilde{\mu} \in (0, 1)$ such that no bonus pay is used ($\sigma = A$) if and only if $\mu \geq \tilde{\mu}$.*

If there are lots of intrinsically motivated agents then the firm will exploit the fact that it can economize on bonus pay and will choose to eschew a pay for performance regime. In principal this logic can explain why some organizations eschew bonus pay. The analysis of Gittleman and Pierce [2013, Table 1] suggests that around 40% of the private sector in the U.S. uses performance related pay. But there are big sectoral differences varying from less than 20% in the hospitality and leisure industries to nearly 70% in financial services. There, are of course, a range of explanations for this. But the prevalence of intrinsic motivation in different organizations is one possible factor.

We now ask what impact the choice of remuneration has on cultural evolution which will depend on the remuneration policy used by the firm. Given the way that we have modeled it, intrinsic motivation is not evolutionarily stable. Specifically:

Proposition 2 *For all* $\mu_0 \in [0,1]$ *the only stable steady state is* $\mu = 0$.

It is crucial for this result that we have not allowed intrinsic motivation to generate either positive or negative utility *per se*. This assumption, combined with motivation crowding-out $\gamma \leq 1$, means that there is no psychological fitness advantage to being an intrinsically motivated type. Thus, even though there could still be intrinsically motivated workers in organizations, we should not expect intrinsic motivation of this kind to survive in the long-run. This would, of course, change this conclusion if there were a pure joy of motivation, paralleling the idea of warm glow in the literature on charitable giving. In that case, there would be a fitness advantage to the intrinsically motivated type.[14] In that case, we could have $\Delta(\mu) > 0$ for $\mu \geq \tilde{\mu}$ and, if Assumption 2 holds, then Lemma 2 would apply creating the possibility of multiple steady states at $\mu = 0$ and $\mu = \bar{\mu}$. Another possibility is that some workers are intrinsically motivated already when they are hired, due to some pre-workforce socialization by parents or schools. Then in spite of the fact that motivation being destroyed by socialization at work, there would be a fresh stock of motivated workers replenishing the workforce and sustaining a limited level of intrinsic motivation over time.[15]

It is worth noting that intrinsic motivation declines independently of whether steady state profits are higher in the long-run when all workers are intrinsically motivated. Thus, it is quite possible that

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

Besley and Ghatak (2016) note that long-term contracting between the firm and works might work to alleviate this problem. Hence as Besley and Persson (2017) observes, this is best thought of as a failure of the Coase theorem for organizations in the presence of cultural evolution since firms cannot credibly commit to reward intrinsically motivated workers.

[14] This could also arise in a model such as Benabou and Tirole (2003) which develops an approach to intrinsic motivation where the principal has some information relevant to agent's payoff and can influence the agent's beliefs about this thereby affecting the agent's effort. In cases where the principal is able to harness an agent's motivation, this will also give them a payoff from undertaking a task which those who are not intrinsically motivated do not receive. This would give a boost to the psychological fitness of intrinsically motivated agents.

[15] In similar vein, Bisin and Verdier (2001) conjecture that a strong hispanic culture prevails in the United States despite strong assimiliation forces due to a continuous inflow of new hispanics.

By making the assumption of no intrinsic benefit to intrinsic motivation, this section provides a useful benchmark for what comes next. The next two models ground motivation in more specifically – the choice of a mission or reciprocity. We show that there are then conditions under which motivated agents survive the cultural dynamic that we are proposing.

4.2 Mission preferences

Now suppose that motivation is linked to a specific action by the firm, namely something which reduces profits but increases the payoff of the motivated agents and hence makes them more inclined to put in effort. We will think of this as a pro-social action but it could also be giving a perk to the workers which only one group of workers values and makes them more willing to work hard for the firm e.g. reducing the carbon footprint of the firm. Following Besley and Ghatak (2005) we will refer to this as mission motivation as the worker prefers to work for a firm which shares their values or organizes production in ways that are congenial for workers. Here, however, we do not allow workers to be heterogeneous in their mission preferences ex ante but for these to evolve over time due to socialization within the firm. We continue to focus on the case where the firm is profit-maximizing rather firms also having mission preferences.[16]

Suppose then that there is an action which costs the firm c but which has value ν to the motivated types. 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 and pastries. In the latter case, it is simply a form of benefit which rewards the work for better performance implicitly by reducing the disutility of effort. We propose the following formulation of preferences to capture this:

$$v\left(m,\sigma,\mu,\psi,\delta\right)=\begin{cases}\delta\left[\nu-\psi\right] & \text{if } \sigma=A\\ -\delta\psi & \text{otherwise}\end{cases}$$

where $\nu>0$ is non-pecuniary benefit for the motivated type. We suppose that $v\left(s,\sigma,\mu,\psi,\delta\right)=-\delta\psi$ for $\sigma\in\{A,B\}$, i.e., the standard formulation of disutility of effort applies to selfish workers.

[16] This rules out the possiblity of sorting among workers and firms to take advantage of mission-motivation.

There is no direct restriction on the choice of bonuses in this case except that $\beta \geq 0$. In particular, even for $\sigma = A$, we will generally have $\beta > 0$. Let $\hat{\beta}_\sigma(\mu)$ once again denoted the optimal remuneration contracts. The profit of a firm which chooses the private action generating a local public good for the workers ($\sigma = A$) is

$$\hat{\Pi}(\mu, A) = \max_{\beta \geq 0} \left\{ [y - \beta] \alpha \left[\mu F(\nu + \alpha\beta) + (1 - \mu) F(\alpha\beta) \right] - c \right\},$$

and with $\sigma = B$, it is

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

Note that in this case $\hat{\beta}_B(\mu)$ is independent of μ. We now have the following result comparing profits across the two choices available to the firm:

Proposition 3 *For small enough c, there exists $\tilde{\mu}$ such that the firm uses pro-social motivation if and only if $\mu \geq \tilde{\mu}$.*

This result is similar to Proposition 1. Firms will choose to reward their motivated agents only when there are sufficiently many of them and it is cheap enough to do so. It is easy to see that, as in Besley and Ghatak (2005), $\hat{\beta}_A(\mu) < \hat{\beta}_B$, i.e. bonuses are lower when the firm makes use of mission-motivation.

Once again we can explore how this affects the cultural dynamics. For this case we have the following result:

Proposition 4 *For all $\mu_0 > \tilde{\mu}$, then if u is small enough and Assumption 2 holds, the organization converges to a steady state where $\mu = \bar{\mu} < 1$ 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.

Hence the model predicts that mission motivation can survive but only above a threshold. This is in contrast with the case of intrinsic motivation where being motivated did not generate any positive utility that could have fitness advantage in cultural evolution. But if the natural state is $\mu_0 = 0$, this is not promising in the case of pure profit maximization. But as we discuss in the extensions section below, a firm set up a founder who has mission preferences and gets utility from a pro-social mission can create an environment for the emergence of a persistent pro-social culture in the organization.

4.3 Reciprocity

We now explore the possibility that agent motivation can be due to a firm offering a "fair" reward structure which delivers equal pay for all agents in the firm. Hence, just as in the case of pure intrinsic motivation, we consider a case where there are no individual bonuses, i.e. $\beta_A(\mu) = 0$ for all μ. We continue to model $\sigma = B$ as a standard incentive arrangement where the firm can set individualized rewards.

Motivated agents will now put in effort when $\sigma = A$. However, we now think of this as based on positive reciprocity, whereby effort is exchanged for the firm offering the same reward to all agents, i.e. not differentiating the pay of those who produce more. However, the expectation on the part of motivated agents is that all of those who work in the firm put in effort. Those who do not follow this norm are punished, a feature which has been observed in lab experiments and is referred to as *negative reciprocity*. Moreover, such punishments are meted out to shirkers even if it is costly to those who make them. To "administer" such punishments, we suppose that agents can observe other agents' effort decisions. Since motivated agents never shirk, the cost of such punishment is $p(1-\mu)\chi$ where χ is the fraction of unmotivated workers who shirk and p is the cost of punishing. We suppose that this generates a punishment for the shirking workers of μP which naturally is increasing in μ, the size of the group of motivated workers.

Given the specific structure of the model, the level of the fixed wage paid to workers reflects their outside option. Hence the system of rewards is

egalitarian but with the benefits accruing to the firm in the form of higher profits. The reciprocity we are focusing applies *between* workers in the firm rather than the workers and the firm's owners. However, analytically, the argument developed in this sub-section holds regardless of the fixed wage policy pursued by the firm when $\sigma = A$. It could, for example, be the case that there is a fair wage-effort relationship in the firm of the kind posited in Akerlof and Yellen (1990) and that even motivated workers shirk when they paid below the fair wage. Here, we are supposing that the fair wage is the outside option which is normalized at zero.

We capture these motivations as follows by supposing that the payoff of motivated agents is:

$$v(m, \sigma, \mu, \psi, \delta) = \begin{cases} -p(1-\mu)\chi & \text{if } \sigma = A \\ -\psi\delta & \text{otherwise.} \end{cases}$$

Although with $\sigma = A$, they do not bear a disutility of effort, they now bear a punishment cost. However, they have a standard disutility of effort when $\sigma = B$, recognizing that positive reciprocity is attached only to the case of egalitarian incentives. For the selfish agents, the payoff is

$$v(s, \sigma, \mu, \psi, \delta) = \begin{cases} -\mu P(1-\delta) - \psi\delta & \text{if } \sigma = A \\ -\psi\delta & \text{otherwise.} \end{cases}$$

In addition to the standard cost of effort, there is now a punishment from shirking inflicted by the motivated agents. We now explore the level of effort, and firm level output, under each of the remuneration schemes.

If $\sigma = A$, all of the motivated put in effort and punish those agents who have shirked so that selfish agents put in effort if and only if $\psi \leq \mu P$,i.e. their private cost of effort is less than the punishment they face. Then the fraction of agents who put in effort, is given by $F(\mu P)$, i.e. the fraction of shirkers among the unmotivated is $\chi(\mu) = 1 - F(\mu P)$. Having more motivated agents in the organization reduces shirking since there is stronger peer-pressure. Thus total output with $\sigma = A$ total output is

$$\tilde{X}(\mu) = \alpha[\mu + (1-\mu)F(\mu P)]$$

which is increasing in μ. There is no inequality of reward when $\sigma = A$. Profits in this case are given by:

$$\hat{\Pi}(A, \mu) = \tilde{X}(\mu) y$$

recalling that $\hat{\beta}_A(\mu) = 0$ by assumption. Thus motivated workers are not offered a share of output, just their fixed wage. As above, having many motivated agents makes offering a purely flat wage optimal. Thus, it provides a micro-foundation for the pure intrinsic motivation case. This is because effort is provided for "free" and motivated agents also use peer pressure to elicit effort from selfish agents.

When $\sigma = B$, then the analysis is exactly as in the past two sections. In this case

$$\hat{\Pi}(B, \mu) = \max_{\beta \geq 0} \{(y - \beta) F(\beta\alpha) \alpha\}$$

Denote the optimal bonus share as $\hat{\beta}_B(\mu)$. Now we can compare the choice of organizational form where we have a result paralleling Propositions 1 and 3 above.

Proposition 5 *There exists $\tilde{\mu}$ such that the firm prefers no individual rewards $\beta_A(\mu) = 0$ if and only if $\mu \geq \tilde{\mu}$.*

This is similar to above where having many motivated agents makes relying on flat incentives optimal for a profit maximizing as it can harness such motivation with minimal rewards.

We can now consider how cultural evolution proceeds in this case. In looking at the utility difference between motivated a selfish agents, bear in mind that peer pressure imposes a utility loss on both groups of agents. For motivated agents, it is costly to punish the selfish agents and the selfish agents are the recipients of those punishments.

We now have the following result showing when this form of reciprocal motivation can survive when firms endogenously choose the structure of rewards based on profit maximizing considerations:

Proposition 6 *Suppose that*

$$-(1 - \tilde{\mu}) \chi(\tilde{\mu}) p + \int_0^{\tilde{\mu}P} \psi dF(\psi) > 0$$

then, if u is small enough and Assumption 2 holds, for all $\mu > \tilde{\mu}$, the firm converges to a steady state where $\mu = \bar{\mu}$ in the long run for $\mu_0 > \tilde{\mu}$. Otherwise, the only stable steady state is $\mu = 0$.

In this case, it is not sufficient that the organization picks the organization that exploits motivated agents at the initial condition. This is because this does not guarantee that motivated agents have a fitness advantage which depends on whether the cost of making punishments by motivated agents exceeds the cost of receiving them by selfish agents. So, even if it is optimal to pick $\sigma = A$ initially, there may be a time path towards reductions in motivation leading to individualized rewards and higher inequality overtime. Hence this model adds an extra consideration which did not arise in the model with pure mission motivation. However, such organizations may be able to sustain this modus operandi when it is not too costly to be the type who uses negative reciprocity to punish the non-cooperative agents. These organizations will preserve this system of rewards.[17]

5 Extensions

We now outline some extensions of the core ideas which draw on the different applications above.

5.1 A Role for Motivated Founders?

We have assumed so far that leader's are purely profit oriented. We now consider what happens if we allow them to care intrinsically about choosing $\sigma = A$ in the applications above. A key issue is how having motivated founders can change the organizational dynamic thereby affecting the evolution of μ. We will explore this for each application in turn.

In the first application where a leader of an organization chooses $\sigma = A$ intrinsically, it is still the case that $\mu = 0$ is the ultimate long-run equilibrium as a psychological fitness disadvantage prevails to being the motivated type. This further illustrates why pure intrinsic motivation has a fragile quality as the basis for running an organization.

[17] Even though there is a preference for fairness, this does not apply to the *level* of rewards paid to all workers by the firm. A concern for fairness in rewards could place a lower bound on b_A so that workers are paid above their subsistence wage. Propositions 5 and 6 would hold for any given level of $b_A \in \left[\frac{z}{\bar{X}(\mu)}, y\right]$. We could interpret $b_A = y$ as a workers' cooperative where there is no residual claimant on profit which is shared among all of the workers. Effort in the cooperative is sustained by reciprocity.

In the second application, having a leader committed to $\sigma = A$ always gives a fitness advantage to the motivated types. Now the organizational will converge to having a motivated workforce with $\mu = \bar{\mu}$ even if it begins from $\mu = 0$. This is interesting since it suggests that a founder who runs the firm in a particular way can have a long-run impact as long the firm is run by the founder to the point where $\mu \geq \tilde{\mu}$ where $\tilde{\mu}$ is defined in Proposition 3 as along as u is small enough. Thereafter, even if the firm is taken over by a pure profit maximizer, it will be run as a motivated firm. A classic example which fits this path is the takeover of Ben and Jerry's Ice Cream by Unilever which led many to doubt whether it would preserve its ethical stance on the sourcing in inputs. But 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. However, if the takeover is too early in a firm's history, then a profit-maximizing leader would revert of $\sigma = B$ and the firm would converge back to $\mu = 0$ as in Proposition 4.

The third application adds an additional consideration. Having a motivated leader committed to $\sigma = A$ need not create an organization with a workforce motivated by fairness if μ is initially sufficiently low since there is no fitness advantage to the fairness motivated types until μ is sufficiently high. This is because peer pressure costs for the few motivated agents at the starting point is insufficient to create a positive psychological fitness advantage to such agents. Building a culture of fairness among the employees will happen when the leader is committed to $\sigma = A$ only if, at the initial condition μ_0,

$$-(1-\mu_0)\,\chi(\mu_0)\,p + \int_0^{\mu_0 P} \psi dF(\psi) > 0 \tag{7}$$

holds. Otherwise, μ will converge to zero even if the founding leader believes in fairness. Moreover, the profits of the firm will be lower than if it was run as a profit maximizing firm. The converse is true in the case where (7) holds. Now the founder will have a permanent effect on the firm once μ is high enough. Hence, even if the leader of the organization is no longer committed to fairness, it will optimal to run the firm this way once the founder is no longer running the firm. This example casts light on why workers' cooperatives are quite rare even the founders are committed to running firms this way. But once the culture is entrenched, such firms will persist.

5.2 The Impact of Regulation

Regulation which affects the firms choice between $\sigma = A$ and $\sigma = B$ will have a similar effect to being run by a motivated leader. Perhaps the most interesting case is where there is mission motivation and regulation forces a firm to pick a specific mission. This will create a fitness advantage for the mission-motivated type which in turn leads μ to grow over time. This may have a persistent effect on the organization in the sense that, if the regulation is abolished, it will not necessarily be optimal for the firm to choose $\sigma = B$ since $\mu \geq \tilde{\mu}$.

To be more concrete, suppose that there is a regulation which forces a firm to have a green technology. This may lead to a temporary loss of efficiency to the firm by lowering profits (due to the cost c). But it will also give a fitness advantage to environmentally motivated workers whose payoffs are now higher than those who do not value the environmental stance of the firm. Eventually the green technology may 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.

5.3 The Evolution of Inequality

Studying motivational dynamics can also provide insights into inequalities in the wage structure. One of the most striking findings in Piketty (2013) is that there was a sharp fall in inequality following the Second World War. One popular view of this is in terms of shifting social norms. Moreover, the observation that this unravelled in the 1980s and onwards is attributed to a norm dynamic which tolerated great inequality.

Our third application can provide some insight into that process since it has a prediction for how inequality in organizations can co-evolve with agent motivation. When $\sigma = A$, there is no differentiation in rewards. However in a type B organization the variance of rewards inside the organization. An obvious measure of inequality in rewards within an organization is:

$$\left(F\left(\hat{\beta}\left(\mu\right)\alpha\right)\alpha\left(1-F\left(\hat{\beta}\left(\mu\right)\alpha\right)\alpha\right)\right).$$

We can then interpret changing inequality over time varying with μ.

If the second world war caused a positive μ shock brought about by socialization during war service then this could reduce inequality as firms

adopt more egalitarian wage structures in line with Proposition 5. However, whether this is sustained over time is not clear as Proposition 6 shows. As we saw, this depends on the effectiveness of the peer pressure mechanism needed to sustain it and the exact initial condition. In some organizations, a high value of μ would be able to sustain continued use of egalitarian rewards whereas in others it could unravel over time. This could explain why inequality rises again over time as fewer workers that were subject to the position μ shock are in the work force and there are insufficiently strong incentives to become a reciprocal type among those who remain.

5.4 Revenue Shocks

Things which affect productivity, i.e. increase output are neutral with respect to the choice of reward structures. However things which make some activities relatively more profitable in the market are not. Specifically, consider what happens when there are increases in the price of a good per unit of output, i.e. y in the model. With a modest strengthening of Assumption 1, all three of our applications imply that this will tend to favor choosing $\sigma = B$ over $\sigma = A$. Formally, we state this as:

Proposition 7 *Suppose that $F(\cdot)$ is concave then an increase in y raises the threshold above which $\sigma = A$ is chosen in all three applications.*

This suggests that non-standard incentives are likely to be used only in firms which are relatively less profitable. In the case of mission motivation and fairness motivation, this has implications for the evolution of motivation since the threshold for μ at which there is convergence to 1 is higher.

5.5 Strategic Socialization

In our model, firm optimize without taking into account the effect that their decisions have on the socialization process. However, it is possible that they understand that they are creating cultural as well as economic incentives. This would potentially affect decisions over β and σ. This may create a trade-off for the firm between which system remuneration is short term versus long-term optimal and the discount factor used could therefore affect the strategy pursued. To provide a benchmark, suppose that the firm is patient and cares only about long run payoffs where $\mu = 1$ or $\mu = 0$. This

would then look a lot like the incentives that we discussed under motivated leaders above. But, as we cautioned in that case, our third application where agents care about reciprocal behavior, there may not be any way to achieve $\mu = 1$ given the punishment technology that workers have.

This discussion assumes that there are no other tools available to the firm to socialize workers. But it could be that training could be used as part of the process, or firms can take overt efforts to project the fortunes of different types using narratives or portraying stereo types. How far this has to be anchored in rational worker behavior is moot. In standard economic models, the assumption is that agents are not systematically fooled and can "see through" efforts by firms to influence them. Attempts to socialize workers may also come from outside the organization. For example, countries with strong ideologies such as communist countries like North Korea engage in intense propaganda for national loyalty. This may have a particular impact when there are is a large state-owned enterprises. Opening the socialization black box is an interesting topic for future research.

6 Concluding Comments

This paper has suggested a framework for studying the evolution of motivation alongside the reward structures in firms. It has emphasized how these co-evolve and that the choice of reward structures can either enhance or diminish intrinsic motivation, mission motivation and fairness motivation. For the latter two, we have shown that organizations can harness non-pecuniary motivations even when the goal of the organization is profit-maximization. However, there are natural threshold effects which mean that this takes hold only when μ is sufficiently high. Otherwise, there is a descent towards standard individualized incentives and 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). But to appreciate their arguments, it is necessary to utilize the idea that preferences are endogenous. Although this remains a debating point in economics, the tools developed, for example, in Alger and Weibull (2013) and Dekel et al (2007) open up these possibilities. Putting structure on this helps to give some discipline to this process and illustrates the limits on the arguments. It also illustrates the range of circumstances in which *homo economicus* has a fitness advantage.

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

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

[6] Besley, Timothy and Torsten Persson, [2017], "The Joint Dynamics of Organizational Culture, Design, and Performance," unpublished typescript.

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

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

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

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

[11] 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.

[12] Bolton, Gary E. and Axel Ockenfels, [2000], "A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90(1), 166-193.

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

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

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

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

[17] 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.

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

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

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

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

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

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

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

[25] 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.

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

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

[30] Kohn, Alfie, [1993], "Why Incentive Plans Cannot Work," *Harvard Business Review*, September-October, 16-21.

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

[32] Levine, David K., [1998], "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics* 1, 593–622.

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

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

[35] Nowak, Martin A., [2006], "Five Rules for the Evolution of Cooperation," *Science*, 314, 1560-1563.

[36] Piketty, Thomas, [2013], *Capital in the 21st Century*, Cambridge AM: Harvard University Press.

[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] Rabin, Matthew, [2003], "Incorporating fairness into games theory and economics," *American Economic Review*, 83(5), 1281-1302.

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

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

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

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

[43] Sobel, Joel, [2005], "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43(2), 392-436.

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

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

[46] Trivers, Robert L. [1971], "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology*, 46(1), 35–57.

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

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

Appendix

Proof of Lemma 1 Assumption 1 ensures that the following problem is well-behaved:

$$\beta \in \arg\max \Pi(\beta)$$

where $\Pi(\beta) = (y - \beta) H(\beta)$ and

$$H(\beta) = \mu F(\nu + \gamma\alpha\beta) + (1 - \mu) F(\alpha\beta)$$

where $\nu > 0$ and $\gamma \leq 1$. There are two applications that we will study, one for which $\nu = 0$ and $\gamma < 1$ and another for which $\nu > 0$ and $\gamma = 1$.

To prove Lemma 1, first note that a standard result is that $\Pi(\beta)$ is quasi-concave if $\log(H(\beta))$ is concave, i.e.

$$\frac{h'(\beta)}{h(\beta)} - \frac{h(\beta)}{H(\beta)} < 0$$

where we use $h(\beta)$ for the derivative of $H(\beta)$, like a density function, and $h'(\beta)$ for its derivative. This can be rewritten as

$$\frac{H(\beta) h'(\beta)}{\{h(\beta)\}^2} < 1. \tag{8}$$

Now, the first-order condition for the choice of β is

$$\Pi'(\beta) = -H(\beta) + [y - \beta] h(\beta) = 0.$$

Evaluated at $\beta = 0$,

$$\Pi'(0) = -\mu F(\nu) + \alpha y \{\gamma\mu f(v) + (1 - \mu) f(0)\}$$

The second-order condition for the choice of β is:

$$-2h(\beta) + [y - \beta] h'(\beta) < 0.$$

Plugging in the first-order condition, the condition becomes:

$$-2h(\beta) + \frac{H(\beta) h'(\beta)}{h(\beta)} < 0$$

or,

$$\frac{H(\beta) h'(\beta)}{\{h(\beta)\}^2} < 2.$$

This holds if (8) holds. So it is sufficient to assume that $\log(\mu F(\nu + \gamma\alpha\beta) + (1 - \mu) F(\alpha\beta))$ is concave.■

Proof of Lemma 2 Recall that

$$\mu_{t+1} - \mu_t = \mu_t \Gamma(\mu_t).$$

Observe that

$$\Gamma(1) = u\left[\rho(1-u)\left\{2G(\Delta(1)) - 1\right\} - 1\right] < 0 \tag{9}$$

as ρ, u, and $G(\Delta(1))$ are all less than or equal to 1. Also,

$$\Gamma(0) = \rho(1-u)\left\{2G(\Delta(0)) - 1\right\} - u.$$

For $G(\Delta(0)) > \frac{1}{2}$, so long as u is small enough, $\Gamma(0) > 0$. Otherwise, for $G(\Delta(0))$ small enough, $\Gamma(0) < 0$.

Now, consider (5). For $u = 0$, we get $\Gamma(\mu) = \rho(1-\mu)\left\{2G(\Delta(\mu)) - 1\right\}$. For $\Delta(\mu) > 0$, $G(\Delta(\mu)) > \frac{1}{2}$ and so $\Gamma(\mu) \geq 0$. Since $\Gamma(\mu)$ is decreasing in u, for a small enough value of u, $\Gamma(\tilde{\mu}) > 0$ assuming there exists $\tilde{\mu}$ such that $\Delta(\tilde{\mu}) > 0$.

Since $\Gamma(\mu) > 0$ is continuous and we assume in Assumption 2 that it is strictly concave (9), then the intermediate value theorem implies that there exists two values of μ, $\left\{\underline{\mu}, \bar{\mu}\right\}$ with $1 > \bar{\mu} > \underline{\mu} \geq 0$ such that $\Gamma\left(\underline{\mu}\right) = \Gamma(\bar{\mu}) = 0$. Moreover, $\Gamma'\left(\underline{\mu}\right) > 0 > \Gamma'(\bar{\mu})$. Hence for all $\mu > \underline{\mu}$ $\mu_{t+1} - \mu_t > 0$ and for $\mu > \mu_H$, $\mu_{t+1} - \mu_t < 0$. So $\bar{\mu}$ is a stable steady-state when $\mu > \underline{\mu}$, i.e. $lim_{t-\infty}\mu_t \to \bar{\mu}$. ■

Proof of Proposition 1 Assumption 1 guarantees that $\hat{\beta}_B(\mu) > 0$ for all $\mu \in [0,1]$. By Lemma 1, this maximization problem is well-behaved and has an interior solution $\hat{\beta}_B(\mu)$. Now note that at $\mu = 0$ we have:

$$\hat{\Pi}(0, A) = 0 < \alpha F\left(\alpha\hat{\beta}_B(0)\right)\left[y - \hat{\beta}_B(0)\right] = \hat{\Pi}(0, B).$$

Also, for $\mu = 1$ we have:

$$\hat{\Pi}(1, A) = y\alpha > \alpha F\left(\gamma\alpha\hat{\beta}_B(1)\right)\left\{y - \hat{\beta}_B(1)\right\} = \hat{\Pi}(1, B).$$

Hence for small enough μ, then $\sigma = A$ with the opposite being the case when $\sigma = B$. Finally note that

$$\frac{d\left[\hat{\Pi}(\mu, A) - \hat{\Pi}(\mu, B)\right]}{d\mu} = \alpha\left[y - \left[F\left(\gamma\alpha\hat{\beta}_B(\mu)\right) - F\left(\alpha\hat{\beta}_B(\mu)\right)\right]\left[y - \hat{\beta}_B(\mu)\right]\right].$$

For $\gamma < 1$, $F\left(\gamma\alpha\hat{\beta}_B(\mu)\right) - F\left(\alpha\hat{\beta}_B(\mu)\right) < 0$, while for $\gamma = 1$, $F\left(\alpha\hat{\beta}_B(\mu)\right) - F\left(\alpha\hat{\beta}_B(\mu)\right) = 0$. Either way, this expression is strictly positive. As $\mu \to 0$, $\hat{\Pi}(0, A) = 0 < \hat{\Pi}(0, B)$ and $\hat{\Pi}(1, A) > \hat{\Pi}(1, B)$, with this condition, there must be a $\tilde{\mu} \in (0, 1)$ such that flat (zero) wages are used ($\sigma = A$) if and only if $\mu \geq \tilde{\mu}$.■

Proof of Proposition 2 First, note that in this application:

$$\Delta(\mu) = \begin{cases} 0 & \text{if } \mu \geq \tilde{\mu} \\ \int_0^{\alpha\gamma\hat{\beta}_B(\mu)} \left[\alpha\gamma\hat{\beta}_B(\mu) - \psi\right] dF(\psi) - \int_0^{\alpha\hat{\beta}_B(\mu)} \left[\alpha\hat{\beta}_B(\mu) - \psi\right] dF(\psi) < 0 & \text{otherwise.} \end{cases}$$

The first row is due to the fact that m-type agents work but incur no disutility (and get no bonus) while s-type workers do not work, do not incur any disutility and get no bonus. Then for all $u > 0$,

$$\mu_{t+1} - \mu_t = -u\mu_t < 0$$

using (5) since $G(0) = 1/2$.

Now with $\mu < \tilde{\mu}$, (5), implies that:

$$\mu_{t+1} - \mu_t = \rho(1 - \mu_t(1 - u))\mu_t(1 - u)[2G(\Delta(\mu_t)) - 1] - u\mu_t < 0$$

since $G(\Delta(\mu)) < 1/2$ for all $\Delta(\mu) < 0$. Hence globally $\mu_{t+1} - \mu_t < 0$ and μ converges to zero.■

Proof of Proposition 3 We have

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

and with $\sigma = B$, it is

$$\hat{\Pi}(\mu, B) = \max_{\beta \geq 0} \{(y - \beta)\alpha F(\alpha\beta)\}.$$

Applying Lemma 1, the maximization problems in $\hat{\Pi}(\mu, A)$ and $\hat{\Pi}(\mu, B)$ are well-defined (in the case of $\hat{\Pi}(\mu, B)$ we simply apply Assumption 1 and Lemma 1 with $\mu = 0$).

Note that $\hat{\Pi}(0,A) = \max_{\beta \geq 0} \{[y-\beta]\,\alpha F(\alpha\beta) - c\}$, and note that $\hat{\Pi}(0,A) = \hat{\Pi}(0,B)$ when $c=0$. Then differentiating $\hat{\Pi}(0,A)$ with respect to c and applying the envelope theorem with respect to c, shows that $\hat{\Pi}(0,A) < \hat{\Pi}(0,B)$ for all $c>0$. Now note that if $c=0$, then for $\mu = 1$

$$\begin{aligned}
\hat{\Pi}(1,A) &= \left[F\left(\nu + \alpha\hat{\beta}_A(1)\right)\alpha\right]\left[y - \hat{\beta}_A(1)\right] \\
&\geq F\left(\nu + \alpha\hat{\beta}_B\right)\alpha\left[y - \hat{\beta}_B\right] \\
&> F\left(\alpha\hat{\beta}_B\right)\alpha\left[y - \hat{\beta}_B\right] = \hat{\Pi}(1,B)
\end{aligned}$$

for $\nu > 0$ by the fact that $F(.)$ is increasing where the first inequality holds since $\hat{\beta}_A(1)$ is the profit maximizing bonus.

Hence there exists a range of $c \in [0,\overline{c}]$ where $\overline{c} > 0$ such that:

$$\begin{aligned}
\hat{\Pi}(1,A) &= F\left(\nu + \alpha\hat{\beta}_A(1)\right)\alpha\left[y - \hat{\beta}_A(1)\right] - c \\
&> \hat{\Pi}(1,B).
\end{aligned}$$

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

$$\frac{d\left[\hat{\Pi}(\mu,A) - \hat{\Pi}(\mu,B)\right]}{d\mu} = \frac{d\hat{\Pi}(\mu,A)}{d\mu} = \left[F\left(\nu + \alpha\hat{\beta}_A(\mu)\right) - F\left(\alpha\hat{\beta}_A(\mu)\right)\right]\alpha\left[y - \hat{\beta}_A(\mu)\right] > 0.$$

Given the values of $\hat{\Pi}(\mu,A)$ and $\hat{\Pi}(\mu,B)$ at $\mu=0$ and $\mu=1$, 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 First, note that in this application:

$$\Delta(\mu) = \begin{cases} \int_0^{\nu+\alpha\hat{\beta}_A(\mu)}\left[\nu + \alpha\hat{\beta}_A(\mu) - \psi\right]dF(\psi) - \int_0^{\alpha\hat{\beta}_A(\mu)}\left[\alpha\hat{\beta}_A(\mu) - \psi\right]dF(\psi) > 0 & \text{if } \mu \geq \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases}$$

Note that for $\mu \geq \tilde{\mu}$, the expression in the first row is increasing in μ. Hence Lemma 2 applies as long as Assumption 2 holds and u is small enough and hence $\lim_{t\to\infty}\mu_t = \bar{\mu}$.

Now suppose that $\mu < \tilde{\mu}$, then, since $G(0) = 1/2$, (5) implies that

$$\mu_{t+1} - \mu_t = -u\mu_t < 0.$$

Hence μ converges to zero. ■

Proof of Proposition 5 We have

$$\hat{\Pi}(\mu, A) = \tilde{X}(\mu) y$$

and

$$\hat{\Pi}(\mu, B) = \max_{\beta \geq 0} \{(y - \beta) F(\beta\alpha) \alpha\}.$$

Note that for $\mu = 0$:

$$\begin{aligned} \hat{\Pi}(0, A) &= 0 \\ &< F\left(\hat{\beta}_B \alpha\right) \alpha \left(y - \hat{\beta}_B\right) = \hat{\Pi}(0, B). \end{aligned}$$

For $\mu = 1$:

$$\begin{aligned} \hat{\Pi}(1, A) &= \alpha y \\ &> F\left(\hat{\beta}_B \alpha\right) \alpha \left[y - \hat{\beta}_B\right] = \hat{\Pi}(1, B). \end{aligned}$$

Finally, as $\hat{\Pi}(A, \mu) = \tilde{X}(\mu) y$ (while $\hat{\Pi}(\mu, B)$ does not depend on μ) and $\tilde{X}(\mu) = \alpha[\mu + (1 - \mu) F(\mu P)]$

$$\frac{d\left[\hat{\Pi}(\mu, A) - \hat{\Pi}(\mu, B)\right]}{d\mu} = [1 - F(\mu P) + (1 - \mu) F'(\mu P) P] \alpha y > 0$$

As this is positive for all μ, it holds for $\mu = \tilde{\mu}$ which is the interior point at which $\hat{\Pi}(\tilde{\mu}, A) = \hat{\Pi}(\tilde{\mu}, B)$, with $\hat{\Pi}(\mu, A) > \hat{\Pi}(\mu, B)$ for $\mu \geq \tilde{\mu}$.■

Proof of Proposition 6 In this case:

$$\Delta(\mu) = \begin{cases} -(1 - \mu) \chi(\mu) p + \int_0^{\mu P} \psi dF(\psi) & \text{if } \mu \geq \tilde{\mu} \\ 0 & \text{otherwise.} \end{cases}$$

The top row is increasing in μ but can be positive or negative. If $\Delta(\tilde{\mu}) > 0$ then Lemma 2 applies as long as u is small enough and Assumption 2 holds. Hence μ_t converges to $\bar{\mu}$. If $\Delta(\tilde{\mu}) \leq 0$, then using (5) we have that

$$\mu_{t+1} - \mu_t = \left[(1 - u)(1 - \mu_t (1 - u)) \left[2G\left(-(1 - \mu_t) \chi(\mu_t) p + \int_0^{\mu_t P} \psi dF(\psi)\right) - 1\right] - u\right] \mu_t < 0$$

since $-(1 - \mu_t) \chi(\mu_t) p + \int_0^{\mu_t P} \psi dF(\psi) < 0$. Hence for all $\mu \leq \tilde{\mu}$, μ converges to 0. Finally, if $\mu \leq \tilde{\mu}$, then, after using (5),we have

$$\mu_{t+1} - \mu_t < -u\mu_t$$

since $\Delta(\mu) = 0$. Then μ converges to zero in this case too. ■

Proof of Proposition 7 In general, let

$$\Omega(\mu) = \hat{\Pi}(A, \mu) - \hat{\Pi}(B, \mu).$$

We have already shown in the proofs of Propositions 1, 3 and 5 that this is increasing in μ. Hence it suffices to show that it is decreasing in y evaluated at $\tilde{\mu}$. We will show this in each application.

First, we look at the intrinsic motivation case.

$$\hat{\Pi}(\mu, A) = y\mu\alpha \tag{10}$$

recognizing that payments to workers are proportional to total output and for $\sigma = B$ it is:

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

This implies that

$$\frac{\partial \Omega(\tilde{\mu})}{\partial y} = \alpha [\tilde{\mu} - [\tilde{\mu} F(\gamma\alpha\beta) + (1 - \tilde{\mu}) F(\alpha\beta)]] < 0$$

since

$$\left[y - \hat{\beta}_B(\tilde{\mu})\right] \left[\alpha \left[\tilde{\mu} F\left(\gamma\alpha\hat{\beta}_B(\tilde{\mu})\right) + (1 - \tilde{\mu}) F\left(\alpha\hat{\beta}_B(\tilde{\mu})\right)\right]\right] = y\tilde{\mu}\alpha$$

implies that $\alpha \left[\tilde{\mu} F\left(\gamma\alpha\hat{\beta}_B(\tilde{\mu})\right) + (1 - \tilde{\mu}) F\left(\alpha\hat{\beta}_B(\tilde{\mu})\right)\right] > \tilde{\mu}\alpha$.

Now consider the mission motivation case

$$\frac{\partial \Omega(\tilde{\mu})}{\partial y} = \alpha \left[\tilde{\mu} F\left(\nu + \alpha\hat{\beta}_A(\tilde{\mu})\right) + (1 - \tilde{\mu}) F\left(\alpha\hat{\beta}_A(\tilde{\mu})\right) - F\left(\alpha\hat{\beta}_B\right)\right].$$

Let X^σ be the level of output in a firm with remuneration package σ. To show that this is negative note that, if $F(\cdot)$, is concave then

$$\frac{\partial X^A}{\partial \beta} < \frac{\partial X^B}{\partial \beta}$$

for all $\mu \in [0, 1]$. Then since

$$\left[y - c - \hat{\beta}_A(\mu)\right] \frac{\partial X^A}{\partial \beta} = X^A \text{ and } \left[y - \hat{\beta}_B\right] \frac{\partial X^B}{\partial \beta} = X^B$$

and

$$\left[y - c - \hat{\beta}_A(\mu)\right] X^A = \left[y - \hat{\beta}_B\right] X^B$$

at $\tilde{\mu}$, we have

$$\frac{\left(X^A\right)^2}{\frac{\partial X^A}{\partial \beta}} = \frac{\left(X^B\right)^2}{\frac{\partial X^B}{\partial \beta}}$$

which implies that $X^A < X^B$. This proves the result.

Finally, consider the case of fairness motivation. In this case

$$\frac{\partial \Omega(\tilde{\mu})}{\partial y} = X^A - F(\beta\alpha)\alpha < 0.$$

To see this, note that at $\tilde{\mu}$

$$yX^A = \left[y - \hat{\beta}_B\right] F\left(\hat{\beta}_B \alpha\right) \alpha$$

which implies that $X^A < F(\beta\alpha)\alpha$. ■