Reforming an Institutional Culture of Corruption: A Model of Motivated Agents and Collective Reputation*

Justin Valasek†

June 11, 2015

Abstract

Recent empirical studies suggest that poor public sector performance in developing nations is due in part to the difficultly of selecting workers whose motivation is aligned with the mission of the institution – in direct contrast to evidence from developed nations, public sector workers tend to be less prosocial. Moreover, contrary to the public sector efficiency-wage argument, empirical evidence from developing countries suggests that motivation is weakly increasing in wages. This paper provides an account for this discrepancy between developed and developing nations by analyzing a model where motivated workers value the collective reputation of their institution, e.g. due to a prosocial signaling motive or identity concerns. The initial insight of the analysis is that there exists both a high-reputation, low-wage equilibrium and a low-reputation, high-wage equilibrium. Importantly, the comparative statics of motivation and wage differ between the equilibria: starting from low-reputation, higher wages *crowd in* motivation, while starting from high-reputation, higher wages *crowd out* motivation. The paper also details the implications of this model for successful reform: taking reputation as the state variable, we show that a non-monotonic wage path is required to achieve a transition to the high-reputation equilibrium–an initial wage increase to crowd in motivated workers, followed by a wage decrease to crowd out non-motivated workers.
Keywords: Corruption, Motivated Workers, Institutional Reform.
JEL Classification Codes: D23, D73, L32.

*Thanks to Rachel Kranton, Steffen Huck, Paul Heidhues, Rajshri Jayaraman, Sebastian Kodrisch, Frank Heinemann, Tom Cunningham, Shanker Satyanath, Miriam Golden, Macartan Humphreys, Suresh Naidu and seminar participants at the Berlin Behavioral Economics Workshop, the MPI for Tax Law and Public Finance, Lund University and Queen's University Belfast.
†WZB Berlin. Contact e-mail: justin.valasek@wzb.eu

1 Introduction

> "Sweden's bureaucracy is one of the most impressive in the world...a tradition of an efficient, non-corrupt bureaucracy with an impressive work ethic." – Johan Norberg

Swedish bureaucrats enjoy a reputation as belonging to one of the internationally best-regarded systems of public administration. This reputation also extends domestically, as illustrated by the quote from the Swedish Cato Institute fellow Johan Norberg. Interestingly, these results are achieved with a workforce that is paid 7-14 percent less than peers in the private sector.[1] In the light of the work on public sector motivation, this wage differential is unsurprising: in an efficient equilibrium where motivated workers select into public sector, wages are lower as these workers are compensated by non-pecuniary payoffs of working for a well-regarded public administration.[2]

This intuition, however, raises the question of why other countries fail to achieve efficiency in the public sector.[3] In fact, the lack of a well-functioning public sector has been highlighted as a key impediment to growth and stability in developing nations (Besley and Persson 2010). Moreover, recent empirical research has shown that prospective public-sector employees in developing nations are weakly less prosocial than their peers (Hanna and Wang (2014) and Banuri and Keefer (2014)), and that higher wages increase motivation (Dal Bó et al. (2013)). Here we offer a framework that reconciles these contrasting findings in developed and developing nations Specifically, we argue that wage and mission are not the only important factors motivated workers consider while choosing employment: the reputation of the public institution may play a role as well.

For example, motivated workers may be attracted to join the bureaucracy in Sweden precisely due its reputation. As argued by Akerlof and Kranton (2005), workers may directly value the identity associated with their job, and will logically seek employment in institutions consistent with their personal identity. In turn, institutional identity is a function of both the mission and the culture of the institution – while

[1] Controlling for observables, de Koning et al. (2013) find an average differential of 7 percent amongst central government workers, and 14 percent for local government; in Sweden, working conditions and social benefits are similar in the private and public sector.

[2] See Francois (2000) and Francois and Vlassopoulos (2008) for an overview, and Besley and Ghatak (2005), and Delfgaauw and Dur (2007) in particular for arguments regarding a low public-sector efficiency wage.

[3] For broader evidence of public-sector efficiency equilibria in developing nations, see Gregg et al. (2011) and Dur and Zoutenbier (2012).

a motivated worker might be attracted a job in a well-regarded bureaucracy, they might be negatively disposed towards working for a police force widely viewed as corrupt. Additionally, the collective reputation of an institution can affect worker choice through the channel of prosocial signaling (a la Bénabou and Tirole (2006), and Ariely et al. (2009)) since the collective reputation, or aggregate behavior, of an institution provides a signal of its employees type.[4]

This paper explores the effect of collective reputation on the sorting of motivated workers in a labor market. Specifically, we consider the case where motivated workers value the collective reputation of the institution they work for – defined as average behavior within the institution (following Tirole (1996)). Framed in the context of corruption, motivated public-sector workers derive positive value from a collective reputation for low corruption, and a negative value from a collective reputation for high corruption (we refer to the example of the public sector and corruption here, but, as we discuss below, the model we analyze is general to other applications and is presented in a neutral frame). Our aim is to show how collective reputations can contribute to the persistence of poor institutional performance, and how a culture of corruption can be reformed using a commonly accessible policy tool–wage.

While we remain agnostic to the precise mechanism, we present the results of the motivated signaling model as a relevant benchmark.[5] To summarize, the model relies on two key assumptions: (i) there exists a motivated type who, all else equal, has a higher productivity in the public sector; and (ii) the motivated type values the collective reputation of the public institution due to reputation or identity concerns.[6]

We first show that the model implies multiple equilibria – both high-corruption equilibria and low-corruption (efficient) equilibria may exist for given parameter values. Generally, a high-corruption equilibrium is characterized by a lower proportion of motivated types in the public sector than the population average, and a low-corruption equilibrium is characterized by a higher proportion of motivated types and a *lower* public sector wage than in the high-corruption equilibrium. The reason a lower public sector wage is maintained in the low corruption equilibrium

[4]Since the collective reputation and workforce composition are correlated, value homophily in the workplace (a la Lazarsfeld and Merton (1954)) provides yet another explanation why workers may value the collective reputation of their workforce.

[5]Therefore, in the benchmark model presented in the main text, the collective reputation simply serves as a signal as to the proportion of types in the public sector. In Appendix B we derive analogous results for a more generalized model that admits consumption benefits.

[6]Non-motivated types may also value the collective reputation of the public institution because of reputation concerns à la Bénabou and Tirole (2011); our analysis assumes that the motivated type places a greater weight on the collective reputation than the non-motivated type.

is that pro-social types are compensated by the collective reputation for low corruption, while the low wage deters non-motivated types from entering the public sector.[7]

However, in contrast to previous analyses, this does not imply that a low-corruption equilibrium can be achieved by simply setting a low public sector wage. The low wage is a feature, rather than a cause, of the low-corruption equilibrium. Instead, the effect of a change in the public sector wage depends on the initial starting point: in the high-corruption equilibrium a decrease in the public sector wage *increases* corruption, while in the low-corruption equilibrium a decrease in the wage *decreases* corruption. The intuition behind this result lies in the fact that, holding corruption constant, following a wage decrease an equal proportion of non-motivated and motivated workers will exit the public sector. Therefore, starting from a case of high corruption, the proportion of non-motivated workers increases with a wage decrease, leading to an increase in average corruption and making the public sector even less attractive to motivated workers. By the same mechanism, however, starting from a case of low corruption, a lower wage decreases corruption. These findings organize the data that suggest a public-sector efficiency equilibrium in developing nations, but the reverse sorting and comparative statics in developing nations.

We formally analyze the problem of transitioning from a high-corruption equilibrium to a low-corruption equilibrium by introducing a dynamic process in which a proportion of workers are replaced in each period, implying a natural minimum rate of turnover in the public sector. The policy tool we consider for enacting a transition is the public-sector wage, which can be changed transparently.[8] Additionally, measures to decrease the level of corruption undertaken by non-motivated workers will leave the signaling motivations of motivated workers unchanged, since the aggregate level of corruption in the public institution still gives a perfect signal of the proportion of types in the workforce.

We then characterize a wage path that induces a transition between a high-corruption equilibrium to a low-corruption equilibrium. We find that such a wage path generally involves an initial increase in the public wage to attract more motivated types

[7] Analogous to the efficiency wages in Handy and Katz (1998), Besley and Ghatak (2005), and Delfgaauw and Dur (2007).

[8] As opposed to additional monitoring, where the potential monitors might themselves be corruptible (Svensson (2005)). Niehaus and Sukhtankar (2013) document a case in which wage increases were not passed on to employees. In this case, however, workers were generally aware of the official wage increase, suggesting that the de facto wage increase is a function of the relative bargaining power of the workers. As long as the marginal bargaining power is greater than zero, some of the wage increase should be passed on to workers.

into the public sector (crowding in motivation), followed by a gradual decrease of the wage to drive non-motivated types out (crowding out non-motivation). The intuition for a non-monotonic wage path follows from the comparative statics outlined above: Starting from a point of high corruption, only increasing the public-sector wage will decrease corruption. Wage increases alone, however, cannot transition to the efficient, low-wage equilibrium. A transition can only be achieved if a tipping-point threshold of corruption can be reached through a wage increase, after which the wage must be gradually decreased to push non-motivated types out of the public sector and transition to the low-corruption, low-wage equilibrium.

Lastly, we emphasize that the framework we analyze is not peculiar to the public sector: to the extent that motivated workers value collective reputation of generic institutions, the model pertains to any firm or institution that would find it beneficial to attract motivated workers. For example, firms may seek to replicate the recruiting advantages of, say, Google, whose reputation as a dynamic and attractive employer stems at least in part from the high quality of its existing workforce; economics departments may seek to recruit PhD students who are motivated to join academia rather than the private sector, and these academically-motivated students may in turn value a reputation for academic placements. Crucially, however, we show that a transition from a low reputation to a high reputation is only generally feasible if motivated workers value the mission of the relevant institution. That is, a tipping-point reputation can only be reached through a wage increase if, given a neutral reputation, motivated workers prefer employment in the institution in question over their outside option, as is the case when motivated worker directly value the social output of a public institution (i.e. mission-contingent payoffs a la Besley and Ghatak (2005)). This finding suggests that transitions are not feasible in generic institutions, and may require that a firm actively invest in, say, corporate social responsibility,[9] or that transitions are only possible for departments at universities with an overall reputation for academic achievement.

1.1 Literature

This paper contributes to the literature on public sector motivation and endogenous norms in institutions, and to the theoretical literature on reforming corruption through public-sector wages.

[9]The management literature suggests that corporations engage in charitable activities for precisely this purpose; see for example Bhattacharya et al. (2008) "Using Corporate Social Responsibility to Win the War for Talent."

In a classic study, Wilson remarks that, given the lack of incentives..."what is surprising is that bureaucrats work at all" (1989). More generally, it has been argued that non-monetary incentives in the workplace play an important role in determining worker's behavior (Dewatripont et al. (1999), Akerlof and Kranton (2000), Akerlof and Kranton (2005), and Prendergast (2008)). A subset of this literature considers motivated motivations in the workplace, and has largely focused on optimal contracting in the presence of a motivated type, given that motivated incentives can be crowded out or distorted by traditional monetary incentive contracts (see Francois and Vlassopoulos (2008) and Prendergast (2008) for an overview). For example, Murdock (2002) shows that a principal has an incentive to pursue projects that give the agent the highest degree of intrinsic motivations to induce higher ex ante effort. Dixit (2002) consider workers with non-pecuniary valuations in a multitask environment, and finds that the optimal incentive contracts are unchanged by worker motivation. In contrast, Sliwka (2007), and Ellingsen and Johannesson (2008) detail how optimal incentive contracts change if the contract functions as a signal of an underlying characteristic of the workplace, and hence can affect worker behavior through channels such as reciprocity and conformity.

Another strand of this literature, in which our paper arguably falls, is concerned with the question of optimal contracting with endogenous worker sorting into the public sector, given the presence of different behavioral types (Francois (2000), Besley and Ghatak (2005), Prendergast (2007), Delfgaauw and Dur (2008), Auriol and Brilon (2014)). These papers highlight that the efficiency wage in the public sector should be low relative to the private sector, as a low wage will disproportionately attract workers with public sector motivation who are compensated by non-pecuniary benefits of public-sector employment.

By considering motivated agents who value the collective reputation of an institution, however, the question is transformed from a problem of static equilibrium selection to a problem of dynamic transition, since the collective reputation functions as a state variable. That is, similar to Tirole (1996), the institution and its current workers are burdened with the legacy of past corruption, which implies that the impact of incentives becomes sensitive to the institution's starting point: higher wages decrease corruption in a high-corruption equilibrium, but increase corruption in a low-corruption equilibrium. Therefore, reforming a culture of corruption requires a more complex approach than simply replicating the incentives of a low-corruption institution.

In the literature on institutional norms, Huck et al. (2012) and Fischer and Huddart

(2008) consider the case of where social norms are endogenous to the composition of worker behavior; in this case, social norms can be changed by the monetary incentives offered by the firm, as they indirectly affect conformity norms. A recent paper which takes a similar approach to this paper is Besley et al. (2014). They consider a dynamic model of tax compliance norms and demonstrate how, as new laws are introduced, endogenous social norms explain a lag in compliance. While our paper studies reform through worker selection rather than through changing an explicit norm of a constant workforce, our results are complementary. Specifically, if a norm of conformity also affects worker behavior, then conformity can reinforce multiple equilibria in our setting. However, if wage reforms lead to a workforce that is less likely to engage in corrupt behavior, then a conformity norm will simply function to amplify the decrease in corruption due to selection.

Lastly, we highlight that our results reconcile the theoretical notion of a low efficiency wage in the public sector (see Gregg et al. (2011) for empirical evidence) with the fact that higher public-sector wages are weakly correlated with lower corruption (Treisman (2000), Van Rijckeghem and Weder (2001), and Di Tella and Schargrodsky (2003)) and the empirical finding of Dal Bó et al. (2013) that the public sector motivation of applicants in Mexico, where corruption is relatively high, is increasing with the offered wage. Both results are consistent with the model we analyze: starting from a point of high-corruption, wages must be increased to decrease corruption to "below-average" levels; only then can wages be reduced (gradually) to reach the low-wage efficiency equilibrium.

2 Static Model

In this section we introduce a simple model that illustrates the relevant results.

Firms and workers

There are two institutions in the market, labeled A and B (e.g. the public sector and/or private firms). The analysis focuses on the collective reputation and workforce composition of institution A, while institution B is conceptualized as an outside option employment in a competitive market, which is available to all workers.

There is a continuum of workers of measure one with a compact index set I. Workers are one of two types: Non-motivated or Motivated. Take $a_i = 1$ if worker i is moti-

vated and $a_i = 0$ if non-motivated; a proportion λ of workers are motivated. Workers each have institution-specific abilities: y_i for institution A and x_i for institution B. For simplicity, we constrain $y_i = 1$, while x_i, is heterogenous and distributed according to a uniform distribution with support $[\underline{x}, \overline{x}]$. That is, all agents have same ability at institution A, but vary in their outside option employment opportunity. Additionally, x_i is uncorrelated with worker motivation.

Take $p_i = 1$ if worker i is employed in institution A, and $p_i = 0$ if i is employed in institution B.

Payoffs

Institution A has a demand for labor of measure ν, and receives the following the profit from each individual it hires:

$$\pi_i^A = \pi y_i + \beta \mathbb{1}(a_i = 1) - w_i$$

Where y_i is ability of worker i, β reflects the higher productivity of motivated workers at institution A, and w_i is the wage paid to worker i. Firm A does not observe π_i^A directly, but aggregate profit $\pi^A = \int_I \pi_j^A p_i$ is publicly observable. Since all workers have the same expected profit, we constrain the wage in institution A to be constant across workers, w^A.

Definition 1 (Collective Reputation)
The collective reputation of institution A is equal to $C = \int_I p_i a_i / \int_I p_i$.

Note that we define the collective reputation as the proportion of non-motivated types in institution A rather than aggregate behavior within A; however, the two are equivalent in our model since types perfectly correlate with behavior. In other words, agents can infer the composition of types within institution A by observing A's aggregate profit (performance).

Firm B receives following the profit from each individual it hires:

$$\pi_j = x_i - w_i$$

Where w_i is the wage paid to worker i. The individual's ability, x_i, is perfectly observed by the private firm. Also, the private market is fully competitive.

Non-motivated workers have a standard linear utility function over own consumption:

$$U_b(w_i) = w_i$$

Where w_i is i's wage.

Motivated workers differ from non-motivated workers in three regards: (1) they are more productive if matched with institution A, (2) they may value the mission of firm A, and hence may receive a direct benefit of employment in firm A (as in Francois (2000) and Besley and Ghatak (2005)), (3) they value the workforce composition (collective reputation) of firm A, e.g. due to type signaling or a direct preference for workplace homogeneity. To reflect (2) and (3), motivated workers have a utility function of the following form:

$$U_a(w_i, C) = w_i + v(C)\mathbb{1}(p_i = 1)$$

Where C is the proportion of non-motivated workers in institution A, and $v(C)$ captures motivated workers payoffs from both collective reputation and mission; $v(\cdot)$ is strictly increasing and concave.

We will distinguish between a "generic" firm and a "mission-oriented" firm. In a generic firm, holding constant wage and reputation, a motivated agent perceives employment in firm B and employment in firm A as equivalent. In terms of the model, this implies that $v(\lambda) = 0$; when the reputation of Firm A and the outside option are equivalent, than the relative intrinsic motivation for a motivated worker to work at firm A is null. This contrasts with the case where Firm A is mission oriented, where motivated agents are directly motivated by the mission, or product, of Firm A (as in Besley and Ghatak 2005).

In our model, given the constant product produced by each worker, mission-motivation simply translates into a constant benefit of working for firm A: holding constant reputation and wage between sectors, motivated workers prefer working at the mission-oriented sector. That is, mission orientation can captured by the following assumption: $v(\lambda) > 0$.

Since we are considering the wage of institution A as a policy tool, it is necessary to specify the framework for employment in institution A when it is over-demanded (i.e. demand for employment is greater than ν). Since all workers are ex-ante identical from A's perspective, workers are randomly selected for employment in institution A from amongst the applicants (note that workers always have the outside option

of $w_i = x_i$).

Formally, workers choose $\hat{p}_i \in \{0, 1\}$ at the beginning of the period, which determines employment according to the following rule:

$$p_{i,t} \begin{cases} = 0 & \text{if } \hat{p}_i = 0 \\ = 1 \text{ w.p. } q & \text{if } \hat{p}_i = 1. \end{cases}$$

Where:

$$q = \min \left\{ 1, \frac{\nu}{\int_I \hat{p}_i} \right\}.$$

Equilibrium

Since information is complete, the equilibrium concept we use is NE. That is, an equilibrium is defined by a set of employment choices, $\{\hat{p}_i\}$, such that given w^A, and C, non-motivated workers set $\hat{p}_i = 1$ iff $w^A \geq x_i$, non-motivated workers set $\hat{p}_i = 1$ iff $U_a(w^A, C, p_i = 1) \geq x_i$, and $C = \int_I \hat{p}_i a_i / \int_I \hat{p}_i$.

3 Analysis of Static Model

In the analysis, we will use the terminology Motivated/Non-Motivated to classify equilibria:

Definition 2 (Collective Reputation Motivated/Non-Motivated)
The collective reputation of institution A is Motivated *if $C > \lambda$ and* Non-Motivated *if $C \leq \lambda$.*

That is, an equilibrium is motivated if a higher proportion of motivated workers are employed in institution A, relative to the population average. As discussed above, it may be reasonable that payoffs associated the collective reputation are null when an institution has a "neutral" reputation ($C = \lambda$). However, $v(\lambda)$ may still be strictly positive when institution A has a prosocial output.

First, we characterize equilibria in the static model in terms of cutoff types x^a and x^b:

Lemma 1 (Cutoff Equilibrium)
Given w^A, equilibrium employment decisions are characterized by $\{x^a, x^b\}$, where $\hat{p}_i = 1$ if and only if $a_i = 1$ and $x_i \leq x^a$ or $a_i = 0$ and $x_i \leq x^b$.

The result follows simply from the monotonicity of $U_a(\cdot)$, $U_b(\cdot)$ in w_i, which implies a single-crossing in x_i for each type. Lemma 1 states that, in equilibrium, conditional on type, individuals with relatively low ability in institution B will select into institution A.

Lemma 1 also allows us to characterize equilibrium by identifying the private-sector abilities of individuals who are indifferent between the public and private sector, and a corresponding reputation. That is, an (interior) equilibrium is defined by $\{x^a, x^b, C\}$ that solve the following system of equations:

$$\begin{aligned} x^a &= w^A + v(C), \\ x^b &= w^A, \\ C &= \frac{\lambda(x^a - \underline{x})}{(1-\lambda)(x^b - \underline{x}) + \lambda(x^a - \underline{x})}. \end{aligned}$$

Note that the proportion of non-motivated workers who set $\hat{p} = 1$ depends only on the wage in institution A; therefore, since we define equilibria given w^A, when convenient we refer to A's reputation as a function of x^a only ($C(x^a)$). Since $v(C(x^a))$ is increasing and concave in x^a, either an interior intersection exists, or a corner equilibrium exits ($\underline{x} > w^A + v(0)$ or $\overline{x} \leq w^A + v(1)$).

We will refer to an equilibrium as market-clearing if $\int_I \hat{p}_i = \nu$. While it is possible that the profit maximizing equilibrium need not coincide with a market-clearing equilibrium, it still provides a meaningful benchmark. The following proposition specifies sufficient conditions for the joint existence of low and high-motivation market-clearing equilibria.

Proposition 1 (Existence of Market-Clearing Equilibria)
(i) If $v(\lambda) = 0$, there exists a high-motivation equilibrium if ν if small enough and $v(1)$ large enough such that $\nu < \lambda(\underline{x} + v(1))$, and there exists a low-motivation equilibrium.
(ii) If $v(\lambda) > 0$, there exists a high-motivation equilibrium, and there exists a low-motivation equilibrium if ν, $v(0)$ small enough such that $\nu < -(1-\lambda)(\underline{x} + v(0))$.

Proposition 1 illustrates that multiple equilibria exist when institution A's demand for labor is relatively small compared to the overall labor market, and when the motivated type places a high valuation on reputation. The formal proof of the Proposition is left for the appendix. However, Figure 1 illustrates equilibria for

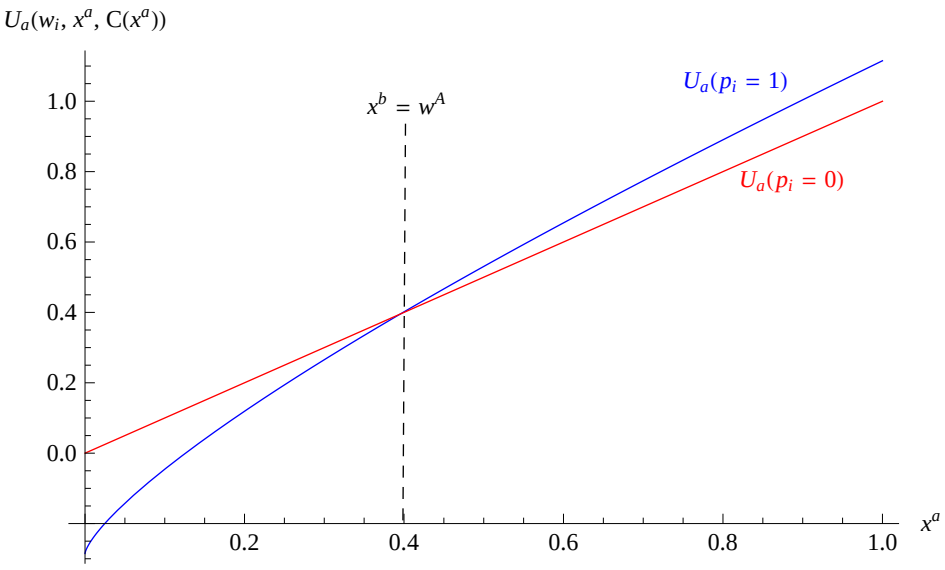

Figure 1: This graph illustrates the respective utility of employment in A and B for a motivated type with $x_i = x^a$, *given* that all motivated workers with $x_i < x^a$ set $\hat{p}_i = 1$. Therefore, given $x^b = w^A$, $C(x^a)$ is increasing with x^a.

a given value of w^A for $v(\lambda) = 0$; the graph illustrates the respective utility of employment in A and B for a motivated type with $x_i = x^a$, *given* that all motivated workers with $x_i < x^a$ set $\hat{p}_i = 1$. Therefore, given $x^b = w^A$, $C(x^a)$ is increasing with x^a. Interior equilibria are represented by intersections of $U_a(p_i = 0)$ and $U_a(p_i = 1)$, since at an intersection, given $C(x^a)$, motivated workers with $x_i = x^a$ are indifferent between employment in institutions A and B.

As illustrated in Figure 1, when $v(\lambda) = 0$, an equilibrium always exists at $x^a = x^b = w^A$, since $C = \lambda$ at this point. This implies that a low-reputation market-clearing equilibrium always exists. However, note that in Figure 1 both corners are also equilibria; that is, $x^a = 0$ and $x^a = 1$ are equilibria since, respectively, $\underline{x} > w^A + v(C(0))$ and $\overline{x} < w^A + v(C(1))$.[10]

Figure 2 illustrates equilibria for a given value of w^A for $v(\lambda) > 0$. The difference between the two cases is illustrated by the fact that $U_a(p_i = 0) < U_a(p_i = 1)$ at $x^a = x^b = w^A$, which shows that $C = \lambda$ is never an equilibrium since if A's reputation matches that of the population average, motivated workers with $x^a = w^A$ strictly prefer employment in institution A. Note that since $U_a(p_i = 0) < U_a(p_i = 1)$ at $x^a = x^b = w^A$, again, either $x^a = 1$ is an equilibrium or $U_a(p_i = 0)$ intersects $U_a(p_i = 1)$ at x^a with $C(x^a) > \lambda$.

[10]Moreover, note that if $v'(\lambda) > 1$, the equilibrium at $C = \lambda$ is unstable, in the sense that the best response dynamics move away from the equilibrium, given a small perturbation from $C = \lambda$.

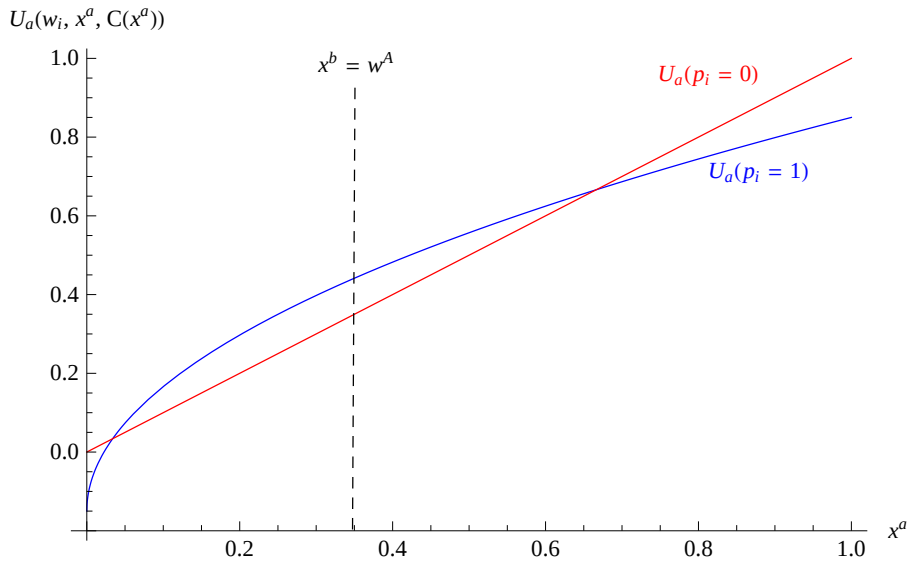

Figure 2: This graph illustrates the respective utility of employment in A and B for a motivated type with $x_i = x^a$, *given* that all motivated workers with $x_i < x^a$ set $\hat{p}_i = 1$. Therefore, given $x^b = w^A$, $C(x^a)$ is increasing with x^a.

Moreover, note that Figures 1 and 2 illustrate an additional result:

Corollary 1

If it exists, the high-motivation equilibrium is unique.

Multiple high-motivation equilibria could only exist if $U_a(p_i = 0)$ intersects $U_a(p_i = 1)$ from below in the range of $C(x^a) > \lambda$; this cannot occur since $v(C(x^a))$ is concave and $v(\lambda) \geq 0$.

The existence of multiple equilibria in this setting is unsurprising – given a homophilous type and two institutions, absent a large wage differential, equilibria will exist where the homophilous type sorts on either institution. Here, however, one equilibrium is superior from the perspective of institution A, since in the market-clearing high-motivation equilibrium the productivity of the average worker is higher *and* the wage is lower. In a static setting, however, it is unclear how institution A can select between equilibria. Therefore, the question we address here is: if institution A finds itself in the low-motivation equilibrium, how can a transition to the efficient, high-motivation equilibrium be achieved? Put differently, how can an institution with a collective reputation of low motivation reform its reputation by inducing high-motivation types to select into the institution?

Addressing this question requires a dynamic version of the model, which is precisely what we introduce in the next section. Additionally, since we are concerned with

the possibility of transition from low to high motivation, for the remainder of the paper we will assume that the parameters of the model are such that there exists a high-motivation market-clearing equilibrium.

4 Dynamic Model

We now add a dynamic layer to the static framework, and consider a discrete-time dynamic framework with an infinite time horizon. A simple repetition of the static game would produce the same equilibria as in the static game – we therefore introduce two logical sources of fiction to the dynamic model.

The first and most important source of friction is that motivated workers value the *lagged* collective reputation of institution A. This captures the notion that reputations and reputation payoffs are sticky, as perceptions might not update automatically (similar to Besley, Jensen and Persson (2014)). Alternatively, this assumption serves as an approximation to a continuous-time model, where only atom-less groups of workers join institution A at any moment. Crucially, this friction implies that transitions from low to high-reputation illustrated in this paper are not achieved through the assumption of coordinated action of a mass of motivated workers.

Formally, workers have period-utility payoffs analogous to the static framework, with the exception that $v(\cdot)$ is a function of (C_{t-1}):

$$U_{a,t}(w_i, C) = w_{i,t} + v(C_{t-1})p_{i,t}$$

Therefore, motivated workers' period payoffs are not a function of their expectations regarding the proportion of motivated workers that will enter firm A's workforce in the *current* period. However, since expectations over *future* periods enter dynamic payoffs, the equilibrium path of $\{C_t\}$ is not independent of expectations.

The second source of friction we introduce is that workers have employment tenure in institution A, in the sense that workers cannot be replaced by A directly. This assumption is of secondary importance (we characterize results when this assumption is non-binding), but increases the verisimilitude of the model to the underlying setting we consider, since it is unlikely that institutions are able to fire and replace all workers in a single period. However, we will clearly detail when and how results are sensitive to this second feature of the dynamic model.

We do incorporate an exogenous method for replacement: a measure $\delta \in (0,1]$ of workers are "replaced" in each period. Workers have an equal probability of being replaced, and are replaced by an individual of the same type and ability ($\{a_i, x_i\}$). Importantly, replaced workers do not have employment tenure ($p_{i,t-1} = 0$ for replacement workers). Therefore, δ both functions as a discount rate, and ensures a minimum level of worker turnover in institution A ($\int_I p_i \delta$). Additionally, workers are always free to exit employment in institution A and take up employment in institution B.

Formally, as before, workers choose $\hat{p}_{i,t} \in \{0,1\}$ at the beginning of each period, and employment according to the following rule that incorporates tenure:

$$p_{i,t} \begin{cases} = 0 & \text{if } \hat{p}_{i,t} = 0 \\ = 1 & \text{if } \hat{p}_{i,t} = 1,\ p_{i,t-1} = 1 \\ = 1 \text{ w.p. } q & \text{if } \hat{p}_{i,t} = 1,\ p_{i,t-1} = 0 \end{cases}$$

Where:

$$q = \min\left\{1, \frac{\delta\nu + \int_I \{\hat{p}_{i,t} = 0, p_{i,t-1} = 1\}}{\int_I \{\hat{p}_{i,t} = 1, p_{i,t-1} = 0\}}\right\}.$$

That is, if the public sector is over-demanded, "open" slots in the public sector ($\delta\nu + \int_I \{\hat{p}_{i,t} = 0, p_{i,t-1} = 1\}$) are randomly allocated to new applicants ($\int_I \{\hat{p}_{i,t} = 1, p_{i,t-1} = 0\}$). Additionally, $C_t = C_{t-1}$ if public-sector employment is zero in time t.[11]

The choice variable of the mechanism designer is a set of wages $\{w_t^A\}_1^\infty$. We assume that the mechanism designer chooses $\{w_t^A\}$ at time zero, and that the the decision is publicly observed. This implies that the mechanism designer has access to commitment; however, we will discuss the sensitivity of the results to no commitment (our main result is robust to no commitment). The timing of the period game is as follows:

1. $\{w_t^A\}$, C_{t-1} observed.
2. Workers choose $\hat{p}_{i,t} \in \{0,1\}$.
3. Period utility ($p_{i,t}$) realizes.
4. δ workers replaced at random.

[11] This rules out transition paths where firm A 'resets' it's collective reputation by choosing a wage low enough such that employment is equal to zero.

Dynamic payoffs

The dynamic setting introduces the possibility of a positive option value of employment in institution A. Therefore, workers' relative utility of employment in A takes the following form:

$$u(w_t^A, C_{t-1}, a_i) - x_i + (1-\delta)O_i^t,$$

where $u(w_t^A, C_{t-1}, a_i)$ is the period t payoff, and O_i^t represents the option value of employment in institution A:

$$O_i^t = (1-\tilde{q}_{t+1})\left[u(w_{t+1}^A, \tilde{C}_t, a_i) - x_i + (1-\delta)O_i^{t+1}\right]$$

Where $\tilde{q}_{t+1}$, $\tilde{C}_t$ is agent i's expectations regarding, respectively, future public sector worker supply and corruption (since expectations must be consistent in equilibrium, we drop the i subscript). Note that the option value at t is non-zero only if $\tilde{q}_{t+1} < 1$; that is, there is no positive option value of holding a public sector job unless the public sector will be over-demanded in the following period. Therefore, O_i^t represents a summation of the benefit of holding a public sector job, relative to applying in the following period, over a contiguous set of periods in which the public sector is over-demanded.

Equilibrium

Since information is complete, the equilibrium concept we use is SPNE. However, since employment decisions are sensitive to future supply for public-sector employment and future workforce composition, we utilize expectations and consistency to identify the equilibria of the model. That is, given $\{w_t^A\}$, an equilibrium constitutes a set of employment choices, $\{\hat{p}_{i,t}\}$, and expectations over the employment decisions of other agents, summarized by expectations over reputation, $\{\tilde{C}_t\}$, and expectations over the number of open slots in the public sector, $\{\tilde{q}_t\}$, such that expectations and choices are consistent, and maximize each worker's dynamic utility:

$$U^t(\{\hat{p}_{i,t}\}, C_{t-1}, a_i, x_i, \{w_t^A\}, \{\tilde{q}_t\}, \{\tilde{C}_t\}).$$

4.1 Analysis of Dynamic Model

First, we look at an individual's decision rule, fixing $\{\tilde{q}_t\}, \{\tilde{C}_t\}$. Fixing expectations, each worker chooses $\hat{p}_{i,t} = 1$ if, and only if, the following expression holds:

$$u(w_t^A, C_{t-1}, a_i) + (1-\delta)O_i^t \geq x_i. \tag{1}$$

Note that the decision rule is independent of $p_{i,t-1}$, since the employment preference is independent of tenure.

Again, given $\{\tilde{q}_t\}, \{\tilde{C}_t\}$, define x_t^a and x_t^b to be the ability of, respectively, the motivated and non-motivated types that are indifferent between working in institutions A and B. That is, x_t^a and x_t^b solve:

$$u(w_t^A, C_{t-1}, a_i) + (1-\delta)O_i^t = x_i.$$

We now characterize an equilibrium in terms of cutoff types x_t^a and x_t^b, analogous to the static case. That is, Lemma 1 extends to the dynamic model since $O_i^t = O_j^t$ for i, j of the same motivation type.

Next, we state a result that will be helpful for characterizing equilibria:

Lemma 2 (Motivated/Non-Motivated Reputation)
Given $\tilde{q}_{t+1} = 1$, A's reputation in period t, C_t, is motivated (non-motivated) if, and only if, $v(C_{t-1}) > 0$ ($v(C_{t-1}) \leq 0$).

Note that an equilibrium is motivated if and only if $x^a > x^b$, and that $O_i^t = 0$ if $\tilde{q}_{t+1} = 1$; therefore, the proof of the lemma follows trivially from the fact that $u(w_t^A, C_{t-1}, a_i = 1) > u(w_t^A, C_{t-1}, a_i = 0)$ iff $v(C_{t-1}) > 0$.

Definition 3 (Steady-State Equilibria)
Given $\{w_t^A\}$ such that $w_t^A = w^A$ for all t, an equilibrium $[\{\hat{p}_{i,t}\}, \{\tilde{q}_t\}, \{\tilde{C}_t\}]$ is a steady-state equilibrium if $\int_I \hat{p}_{i,t} = \nu$, $\tilde{q}_t = \nu$, and $\tilde{C}_t = C$ for all t.

Note that the definition incorporates market-clearing. The relationship between static and dynamic equilibrium is clarified by the following Lemma:

Lemma 3 (Static Equilibrium → Steady-State Equilibrium)
For each market-clearing static equilibrium, there exists a corresponding steady-state equilibrium.

Lemma 3 shows that when there exist both high and low-reputation equilibria in the static model, then there exists corresponding high and low-reputation steady-state equilibria in the dynamic model. The following section will analyze the possibility of a dynamic transition from a low-reputation steady-state to a high-reputation steady-state, precipitated by a designer who controls the wage in institution A.

4.2 Dynamically Strategic Institution

Here we consider a dynamically strategic institution, which sets $\{w_t^A\}_t^\infty$ to maximize their profit stream. Specifically, the objective function of institution A is to chose $\{w_t^A\}$ to maximize $\sum^t (1-\delta^A)^t \pi_t^A$. We consider the situation where A "inherits" a reputation and workforce; that is, institution A is endowed with reputation C_0, and a $t=0$ workforce such that $\int_I p_{i,0} = \nu$ and $p_{i,0} = 1$ iff $a_i = 1, x_i < x^a$ and $p_{i,0} = 1$ iff $a_i = 0, x_i < x^b$. Since we are interested in a transition from a low to high reputation, we constrain $C_0 < \lambda$.

We assume δ^A is low enough that a transition from a low to high reputation is always profitable. That is, since both wage and output is higher in the high motivation steady-state equilibrium, there exists a δ^A low enough so that a transition is profitable, even if it comes at a short-term cost. Moreover, while the analysis is general, we often refer to the case where C_0 corresponds to a low-reputation market-clearing stable point $(\bar{C}_0, \bar{w}_0^A)$, since we place an emphasis on characterizing a transition from a low-reputation steady state equilibrium to a high-reputation steady state equilibrium. In these cases, we refer to $\bar{w}_0^A$ as the starting wage, and our description of wage-path includes $\bar{w}_0^A$.

Formally, we address the question of whether an institution who sets wage path $\{w_t^A\}_1^\infty$ can induce a transition in the state variable, C_t, from $C_0 < \lambda$ to $\bar{C} > \lambda$. Formally, we define a transition in the state variable as follows:

Definition 4
A wage path, $\{w_t^A\}$, transitions from C_0 to C' if, given $\{w_t^A\}$ and C_0, for all equilibria $[\{\hat{p}_{i,t}\}, \{\tilde{C}_t\}, \{\tilde{q}_t\}]'$ either $C_t = C'$ for some t, or *$C_t \to C'$ as $t \to \infty$.*

Note that we do not explicitly seek the wage path that maximized the present value of profits; however, if a transition from low to high reputation is possible, then the profit-maximizing wage path will always transition since wages are lower and output higher in a high-reputation equilibrium.

4.2.1 Example: the case of $\delta = 1$

For expositional reasons, we begin by characterizing transitional wage paths and establishing conditions for their existence given $\delta = 1$. With $\delta = 1$ there is full replacement in each period, and agents' dynamic payoffs are equal to their period payoffs. This allows us to illustrate three central findings of the model in a relatively simple manner.

These main findings are: (1) The relationship between the current-period wage and worker composition is a function of last-period reputation; if $v(C_{t-1}) > 0$, then higher wages crowd out motivated workers, and if $v(C_{t-1}) < 0$, then higher wages *crowd in* motivated workers. (2) If a wage path exists that transitions from an initial low-motivation reputation to the high-motivation steady-state equilibrium, then it is non-monotonic; that is, the wage path involves an initial wage increase followed by a series of wage decreases. (3) A wage path exists that transitions from an initial low-motivation reputation to the high-motivation steady-state equilibrium exists if and only if institution A is mission-oriented; that is, it exists if and only if $v(\lambda) > 0$.

In the following subsection, we generalize the results to interior values of δ and characterize additional results that are peculiar to the general model where dynamic payoffs are relevant and employment tenure can be binding, implying that the reputation of institution A can only be changed gradually.

Formally, given $\delta = 1$, the probability of employment at institution A at time t is independent of employment in period $t-1$ for all agents. That is:

$$p_{i,t} \begin{cases} = 0 & \text{if } \hat{p}_{i,t} = 0 \\ = 1 \text{ w.p. } q_t & \text{if } \hat{p}_{i,t} = 1, \end{cases}$$

where $q_t = \min\{1, \int_I \hat{p}_{i,t}/\nu\}$. Also, since there is full replacement in the public sector in each period, there is no option value of employment in institution A, and workers simply choose $\hat{p}_{i,t}$ to maximize period utility:

$$U^t(p_{i,t}, C_{t-1}, a_i, x_i, w_t^A, \{\tilde{q}_t\}) \begin{cases} = x_i & \text{if } \hat{p}_{i,t} = 0 \\ = q_t(w_t^A + a_i v(C_{t-1})) + (1-q_t)x_i & \text{if } \hat{p}_{i,t} = 1, \end{cases}$$

which implies that workers will maximize their objective using the following simple decision rule:

$$\hat{p}_{i,t} = 1 \textit{ iff } w_t^A + a_i v(C_{t-1}) > x_i.$$

Lastly, note that the following expression characterizes C_t:

$$C_t = \frac{\int_I \hat{p}_{i,t} a_i}{\int_I \hat{p}_{i,t}}.$$

That is, since $\delta = 1$, C_t is simply determined by the current-period employment decisions.

The employment decision rule and the expression for C_t allow us to characterize the relationship between the wage in institution A and its reputation in the current period as a function of its previous-period reputation.

Lemma 4 (Crowding out/in motivation)
If $v(C_{t-1}) \leq 0$, then $\partial C_t(w_t^A)/\partial w_t^A \geq 0$.
If $v(C_{t-1}) > 0$, then $\partial C_t(w_t^A)/\partial w_t^A \leq 0$.

Lemma 4 states the sign of the relationship between current-period wage and reputation depends on whether or not the reputation payoff the motivated type receives from employment in institution A is positive or negative: if the reputation payoff is positive, then higher wages crowd out motivated types; if the reputation payoff is negative, then higher wages *crowd in* motivated types.

The proof follows from the linearity of utility in the public sector wage (formal proof in Appendix). Intuitively, quasilinear utility implies that x^a and x^b are linear functions of w_t^A, which means that a wage increase moves A's reputation closer to λ since it effectively adds a mass of workers to institution A to who have an average motivation equal to the population average. And since $x^a \lesseqgtr x^b$ is determined by $v(C_{t-1}) \lesseqgtr 0$, if $v(C_{t-1}) < 0$ then $x^a \leq x^b$, and therefore an increase in the wage increases the current-period reputation; the analogous argument holds for $v(C_{t-1}) > 0$.[12]

Lemma 4 also provides insight regarding potential transition paths, $\{w_t^A\}$, between an initial, stable point $\{w_0^A, C_0\}$ with low reputation ($C_0 \leq \lambda$), and a stable point, $\{w^{A*}, C^*\}$, with high reputation ($C^* > \lambda$):

Corollary 2 (Non-Monotonic Transition)
Given an initial, stable point $\{w_0^A, C_0\}$ with $C_0 \leq \lambda$, monotonic wage paths do not result in a transition to $C^* > \lambda$:

[12] Simply put, Lemma 4 states that wage increases move the current-period reputation closer to that of the population average. Clearly this will not always be true locally for all distributions of x_i, however, in the extensions, we show that this is true in a neighborhood of $x^a = x^b$, which will allow us to extend the results below to a general distribution of ability types.

1. *For* $\{w_t^A\}$ *s.t.* $w_{t+1}^A \leq w_t^A$, $C_t < C^*$ *for all* t.
2. *For* $\{w_t^A\}$ *s.t.* $w_{t+1}^A \geq w_t^A$, $C_t < C^*$ *for all* t.

Corollary 2 shows that a transition cannot be achieved by wage paths that simply increase or decrease the wage paid by institution A. Therefore, a transition path from $\{w_0^A, C_0\}$ to $\{w^{A*}, C^*\}$, if it exists, must be non-monotonic.

The next result details when a wage path exists that transitions between a non-motivated and motivated reputation, and characterizes the non-monotonic wage path that enables this transition.

Proposition 2 (Existence of Transition)
Given $\delta = 1$, *a wage path that transitions from* $v(C_0) < 0$ *(*$C_0 < \lambda$*) and* $\{w^{A*}, C^*\}$ *exists if, and only if,* $v(\lambda) > 0$.

Proposition 2 demonstrates that the ability to transition from a low to a high reputation depends crucially on whether A is a generic or mission-oriented institution. Existence follows the constructed example below, the proof of non-existence can be found in the appendix.

Non-monotonic transition: Take $v(\lambda) > 0$. The following wage path transitions from $v(C_0) < 0$ ($C_0 < \lambda$) to $\{w^{A*}, C^*\}$ with $C^* > \lambda$:

1. w_1^A solves $w_1^A + v(C_0) = \overline{x}$; that is, w_1^A is set high enough that $\hat{p}_{i,1} = 1$ for all i.
2. w_t^A for $t > 1$ solves $\int_I \hat{p}_{i,t} = \nu$; that is, after period 1, the wage is set at the market-clearing level.

To see why this wage path results in a transition, it is informative to solve for the reputation of Firm A in each period. Initially, the Firm A is endowed with low enough reputation such that $v(C_0) < 0$, and by Lemma 4, A's reputation can only be increased by a wage increase. Taken to the extreme, w_1^A is set at a high enough level such all workers prefer institution A, and A's reputation in period 1 will replicate the population average ($C_1 = \lambda$).

In period 2, reputation is decreasing in wages by Lemma 4 since $v(C_1 = \lambda) > 0$. Therefore, since w_2^A is decreased to the market clearing level, A's reputation will increase ($C_2 > C_1 = \lambda$; the relationship is strict since $v(C_1) > 0$ implies $x^a > x^b$).

In period 3, the market-clearing wage, w_3^A, is lower than in period 2, since em-

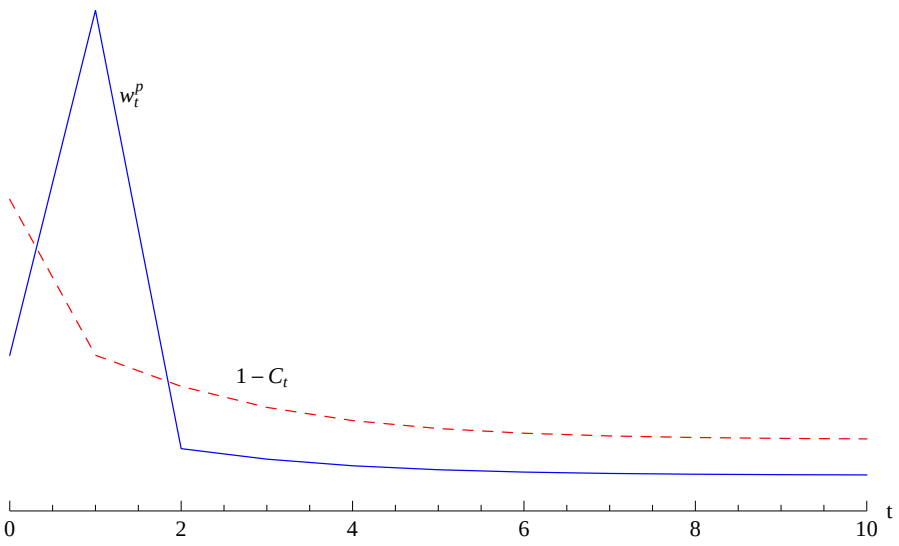

Figure 3: This graph illustrates a wage path that transitions from a high-corruption to a low-corruption equilibrium. Note the initial increase in the public-sector wage (w_t^A; solid line), followed by a decrease and convergence to the efficiency wage. Corruption ($(1 - C_t)$; dashed line), however, decreases monotonically.

ployment in institution A is relatively more attractive for motivated workers given $C_2 > C_1$. By Lemma 4 this implies that $C_3 > C_2$, as the lower wage causes non-motivated workers to exit institution A. By the same logic, in all future periods, the market-clearing wage is decreasing and A's reputation is increasing, implying that $\{w_t^A, C_t\} \to \{w^{A*}, C^*\}$ as $n \to \infty$.

The transition outlined above illustrates the general shape of the non-monotonic path of wages (also illustrated visually in Figure 3). Starting from a low-motivation starting point, w^A must be increased to induce motivated workers to join institution A, hence "purchasing" a higher reputation for motivation. Once a sufficiently high reputation has been reached (with $\delta = 1$ this occurs in a single period), the process is reversed, and public-sector wages are lowered, disproportionately driving non-motivated workers out of the public sector.

Note that this is not the unique transition path, but it ensures full employment during the transition. Other wage paths can converge to $\{w^{A*}, C^*\}$ in finite time: C_2 can be set arbitrarily high by decreasing w_2^A below the market-clearing level. Therefore, there exists a $w_2^{A'}$ such that $C_2 = C^*$, and the high-reputation stable point is reached in period 3. In the following section, we discuss optimal transitions and issues of robustness.

The intuition for the nonexistence result for $v(\lambda) = 0$ follows from the same wage path illustrated above. $C_1 = \lambda$ can always be achieved by setting a high wage in the first period. However, in the second period, an increase in the proportion of motivated workers cannot be achieved through a wage decrease since, by Lemma 4, A's reputation is weakly increasing in w^A.

Proposition 2 and the example transition path demonstrate that the existence of a transition path depends on whether a point such that $v(C_{t-1}) > 0$ can be reached through a wage increase. If not, the region of C where the proportion of motivated workers is increasing in w^A cannot be reached, and a transitional wage path does not exist. In a generic institution, $v(C_{t-1}) > 0$ only if $C_{t-1} < \lambda$, but starting from $v(C_0) < 0$ a point with $v(C_{t-1}) > 0$ cannot be reached through a wage increase. In a mission-oriented institution, however, since motivated workers prefer working in the institution even given a neutral reputation, which implies that $v(C_{t-1}) > 0$ can be achieved through a wage increase, which enables a transition that is unavailable to generic firms.

4.2.2 General analysis: $\delta \in (0, 1)$

The intuition from the example with $\delta = 1$ largely carries over to the more general model. In particular, the following proposition partially characterizes the existence of transitions from a high-corruption to a low corruption equilibrium:

Proposition 3 (Existence of transition $v(C_0) < 0 \Rightarrow \{w^{A*}, C^*\}$)
(i) If $v(\lambda) = 0$ and $v(C_0) < 0$, then for any $\{w^A\}$ there exists an *equilibrium such that $C_t \leq \lambda$ for all t (no transition).*
(ii) If $v(\lambda) > 0$ and $v(C_0) < 0$, then there exists $\{w_t^A\}'$, t' such that $w_t^A \to w^{A*}$ and in all *equilibria and $C_t \to C^* + \epsilon$ for some $\epsilon \geq 0$. (transition in any equilibrium).*

Proposition 3 shows that a transition may not be feasible for a generic firm, but it always possible for a mission-oriented institution. The main difference between this result and Proposition 2 is that Proposition 3 does not fully rule out the possibility of a transition when $v(\lambda) = 0$. Transitions in this case can be achieved in equilibria where are "optimistic" about the future reputation of institution A, and expect A to be over-demanded ($q_{t+1} < 1$). These expectations can be self-fulfilling since they induce a disproportionate number of motivated workers to set $\hat{p}_i = 1$ in the current period to due to the positive option value of employment in A in the following period.

For (ii), first note that an equilibrium exists that transitions to $\{w^{A*}, C^*\}$ that is analogous to the example given for $\delta = 1$. Unlike the case with full replacement, however, since $\delta < 1$, a shift to $C_t = \lambda$ cannot be achieved in a single period: given w^A high enough that all workers set $\hat{p}_i = 1$, A's reputation will increase slowly as only a measure of $\delta\nu$ tenured workers in A are replaced in each period by workers with a higher average level of motivation. However, after C_{t-1} reaches a threshold level where $v(C_{t-1}) > 0$ (this level can always be reached through a wage increase since $C_t \to \lambda$ and $v(\lambda) > 0$), then the transition to $\{w^{A*}, C^*\}$ can be achieved by a market-clearing wage path.

This result merely illustrates that there exists a wage path and a corresponding equilibrium that transitions – multiple equilibria may exist for any wage path. However, for the wage path described above, over $\{w_t^A\}_0^{t'-1}$, where t' is defined as the first period where $v(C_{t-1}) > 0$, the equilibrium is unique since all workers set $\hat{p}_i = 1$. For $\{w_t^A\}_{t'}^{\infty}$ multiple equilibrium outcomes are possible, but since x^b is unique, other equilibria can only occur when x^a is greater than the market-clearing level (other equilibria can occur only if institution A is over-demanded). This implies that in all equilibria, $C_t \geq C'_t$, where C'_t is the reputation in the market-clearing equilibrium.

Example: Expectations-Driven Transition In contrast to the case of full replacement, Proposition 3 does not rule out the possibility of a transition when $v(\lambda) = 0$. Here we explore the conditions under which an expectations-driven transition can be achieved.

To illustrate the possibility of a expectations-driven transition, take the following example: Assume for simplicity that $C_0 = \lambda$ (the population average can always be replicated through a wage increase). The mechanism designer *commits* to the following wage path:

1. $w_1^A = w_2^A$ market-clearing given $C = \lambda$, $O_i^t = 0$.
2. w_t^A market-clearing, given expections that $\int_I \hat{p}_i = \nu$.

Now, suppose workers hold the belief that $C_1 > \lambda$, and hence expect that institution A will be over-demanded in period 2. In this case, $O_i^1(a_i = 1, x'_i) > O_i^1(a_i = 0, x'_i)$. This in turn implies that $v(\lambda) + O_i^1(a_i = 1) > O_i^1(a_i = 0)$, and $C_1 > \lambda$.

That is, expecting that $C_1 > \lambda$ and that institution A will be over-demanded in period 2, the option value of holding a job in A in period 1 is higher for motivated types. This implies that motivated workers will disproportionately enter into institution A in period 1, making the belief that $C_1 > \lambda$ self-fulfilling. After period 2,

given that $C_t > \lambda$ the wage path will transition to a low-corruption equilibrium by the same logic as the proof of Proposition 3.

Note that an expectations-driven transition requires both that workers hold "optimistic" beliefs regarding future levels of corruption, *and* that institution A is able to commit to holding wages above a market-clearing level even after its reputation has increased above λ. Absent commitment, A would prefer to set wages at a market-clearing level in period 2; however, this would imply that $O_i^1 = 0$, which would destroy the incentive for motivated workers to disproportionately enter institution A in period 1. That is, absent commitment to future wages, the expectations-driven transition would unravel.

5 Conclusion

We conclude with some brief comments on the application of our results to reforming an institutional culture of corruption in public-sector institutions. The mechanism for reform that we have highlighted here relies on transforming an institution's reputation through the selection of motivated individuals. Transition, therefore, is likely to be a long-term process, since collective reputations may be slow to adjust to current behavior, and since selection likely requires more time than adjusting the behavior of an existing workforce. Returning to the example of Sweden, in the mid-1800's corruption was endemic to the public administration (see Rothstein (2011) and Sundell (2013)); it was only after a period of transition, involving a radical transformation of the system of payment, that the Swedish bureaucracy evolved into the efficient institution we see today.

Therefore, the most important insight from the mechanism we introduce here might be that the effect of wages on motivation is contingent on the existing composition of workers in the public institution: With a collective reputation for high motivation, higher wages crowd out motivation as per the usual arguments. However, with a collective reputation for low motivation, higher wages crowd in motivation as motivated types require additional compensation for employment in the low reputation institution.

Next, we argue that the mechanism we introduce here is complementary to other efforts at reform. First, it is important to note that the non-pecuniary motives that we analyze depend of the composition of types in the public institution, rather than the precise level of corruption. That is, type-signaling and homophily are

independent of the precise behavior of non-motivated and motivated types, as long as there is a difference in behavior between the two types that can be identified through the aggregate behavior of the institution. Therefore, a direct anti-corruption measure, such as improved monitoring, is orthogonal to the mechanism we present as long as workers update their expectations of each type's behavior.

Second, our mechanism is complementary to efforts to change institutional culture by changing institutional norms: If some proportion of workers are conformist, and hence switch from non-motivated to motivated given some threshold level of aggregate motivation, then increasing the proportion of motivated types in the institution due to selection will precipitate a complementary shift in behavior of the conformist types. This will in turn speed the transition by improving the institution's collective reputation.

Lastly, regarding the robustness of the section of the transition where wages are decreasing in the public institution, note that the transition detailed in the analysis above simply implies that the institution sets a market-clearing wage – insuring that the public institution's demand for labor is met in each period. Theoretically, however, this transition path might not be optimal from the institution's perspective: With a market-clearing wage, the transition to a high-motivation stable equilibrium is achieved through a slow convergence. However, as soon as the institution achieves a high-motivation reputation, the current-period reputation is decreasing in the wage. Therefore, it may be profit maximizing to converge to the high-motivation equilibrium in a single period by slashing wages below the market-clearing level, forgoing profit in the current period, but increasing the current-period reputation and hence increasing profits in the intermediate range.

This theoretical result, however, relies on the assumption that there are no transaction costs involved in switching from the public to the private sector. A more realistic model might include such a transaction cost that is increasing with worker tenure (e.g. due to depreciation of workers' fungible human capital or tenure-based promotion). In this case, a drastic short-term public sector wage cut would disproportionately cause workers with shorter tenure to exit – which would imply a disproportionate exit of motivated workers, since more recent cohorts will have higher average levels of motivation. This mechanism suggests that a drastic wage cut could cause the public institution's reputation to slide back to low-motivation. Therefore, a slow transition that functions predominately through replacing natural turnover with high-motivation cohorts may be more advisable than a temporary and sudden wage cut.

References

Akerlof, George and Rachel Kranton (2000), "Economics and Identity." *The Quarterly Journal of Economics*, 115, 715–753.

Akerlof, George A and Rachel E Kranton (2005), "Identity and the Economics of Organizations." *The Journal of Economic Perspectives*, 19, 9–32.

Ariely, D, A Bracha, and S Meier (2009), "JSTOR: The American Economic Review, Vol. 99, No. 1 (Mar., 2009), pp. 544-555." *The American Economic Review.*

Auriol, Emmanuelle and Stefanie Brilon (2014), "Anti-social behavior in profit and nonprofit organizations." *Journal of Public Economics*, 117, 149–161.

Banuri, Sheheryar and Paul Keefer (2014), "Pro-social behavior where we least expect it? The selection and socialization of intrinsically-motivated government (tax!) officials." *Working Paper.*

Bénabou, Roland and Jean Tirole (2006), "Incentives and Prosocial Behavior." *The American Economic Review*, 96, 1652–1678.

Bénabou, Roland and Jean Tirole (2011), "Identity, Morals, and Taboos: Beliefs as Assets." *The Quarterly Journal of Economics*, 126, 805–855.

Besley, T, A Jensen, and T Persson (2014), "Norms, Enforcement, and Tax Evasion." *Unpublished Manuscript.*

Besley, Timothy and Maitreesh Ghatak (2005), "Competition and Incentives with Motivated Agents." *The American Economic Review*, 95, 616–636.

Dal Bó, Ernesto, Frederico Finan, and Martín Rossi (2013), "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128, 1169–1218.

Delfgaauw, Josse and Robert Dur (2007), "Signaling and screening of workers' motivation." *Journal of Economic Behavior & Organization*, 62, 605–624.

Delfgaauw, Josse and Robert Dur (2008), "Incentives and Workers' Motivation in the Public Sector." *The Economic Journal*, 118, 171–191.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole (1999), "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *... Review of Economic ....*

Di Tella, Rafael and Ernesto Schargrodsky (2003), "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires*." *J LAW ECON*, 46, 269–292.

Dixit, A (2002), "Incentive contracts for faith-based organizations to deliver social services." *Economic theory in a changing world: Policy modelling ....*

Dur, Robert and Robin Zoutenbier (2012), "Intrinsic Motivations of Public Sector Employees: Evidence for Germany ." *Discussion paper / Tinbergen Institute*, 1–32.

Ellingsen, Tore and Magnus Johannesson (2008), "Pride and Prejudice: The Human Side of Incentive Theory." *The American Economic Review*, 98, 990–1008.

Fischer, Paul and Steven Huddart (2008), "Optimal Contracting with Endogenous Social Norms." *The American Economic Review*, 98, 1459–1475.

Francois, Patrick (2000), "'Public service motivation' as an argument for government provision." *Journal of Public Economics*, 78, 275–299.

Francois, Patrick and M Vlassopoulos (2008), "Pro-social Motivation and the Delivery of Social Services." *CESifo Economic Studies*, 54, 22–54.

Gregg, Paul, Paul A Grout, Anita Ratcliffe, Sarah Smith, and Frank Windmeijer (2011), "How important is pro-social behaviour in the delivery of public services?" *Journal of Public Economics*, 95, 758–766.

Handy, Femida and Eliakim Katz (1998), "The Wage Differential between Nonprofit Institutions and Corporations: Getting More by Paying Less?" *Journal of Comparative Economics*, 26, 246–261.

Huck, Steffen, Dorothea Kübler, and Jörgen Weibull (2012), "Social norms and economic incentives in firms." *Journal of Economic Behavior & Organization*, 83, 173–185.

Murdock, Kevin (2002), "Intrinsic Motivation and Optimal Incentive Contracts." *The RAND Journal of Economics*, 33, 650–671.

Niehaus, P and S Sukhtankar (2013), "The marginal rate of corruption in public programs: Evidence from India." *Journal of Public Economics*.

Prendergast, Canice (2007), "The Motivation and Bias of Bureaucrats." *The American Economic Review*, 97, 180–196.

Prendergast, Canice (2008), "Intrinsic Motivation and Incentives." *The American Economic Review*, 98, 201–205.

Sliwka, Dirk (2007), "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *The American Economic Review*, 97, 999–1012.

Svensson, Jakob (2005), "Eight Questions about Corruption." *The Journal of Economic Perspectives*, 19, 19–42.

Tirole, Jean (1996), "A theory of collective reputations (with applications to the persistence of corruption and to firm quality)." *The Review of Economic Studies*, 63, 1–22.

Treisman, Daniel (2000), "The causes of corruption: a cross-national study." *Journal of Public Economics*, 76, 399–457.

Van Rijckeghem, Caroline and Beatrice Weder (2001), "Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much?" *Journal of Development Economics*, 65, 307–331.

6 Appendix A: Proofs

Proof of Proposition 1: We prove the proposition case-by-case:
Case (i), high-motivation: Given $\nu < \lambda(\underline{x} + v(1))$, a market-clearing equilibrium exists where only motivated individuals select into institution A. For $w^A \in (\underline{x} - v(1), \underline{x})$, $x^b = \underline{x}$ in all equilibria since $w^A < \underline{x}$. However, a high-motivated equilibrium exists with $x^a = w^A + v(1)$, where $C = 1$ since all non-motivated workers select into B. Moreover, over $w^A \in (\underline{x} - v(1), \underline{x})$, $\int_I p_i$ is increasing continuously from 0 to $\lambda(\underline{x} + v(1))$ in this high-motivation equilibrium, which implies a market-clearing equilibrium exists since $\nu < \lambda(\underline{x} + v(1))$.

Case (i), low-motivation: Note that if $v(\lambda) = 0$ then a crossing of $U_a(w^A, C(x^a), p_i = 1)$ and $U_a(w^A, C(x^a), p_i = 0)$ exists at $x^a = x^b = w^A$ for $w^A \in (\underline{x}, \overline{x})$. Therefore, a low-motivation market-clearing equilibrium exists at $w^A = \underline{x} + \nu$.

Case (ii), high-motivation: First, we show we show that a high-motivation exists for all values of $w^A \in (\underline{x} - v(1), \overline{x})$. For $w^A \in (\underline{x} - v(1), \underline{x}]$, a unique high-motivation equilibrium exists by the argument in Case (i). For $w^A \in (\underline{x}, \overline{x})$, $v(\lambda) > 0$ implies that $U_a(C(x^a), p_i = 1) > U_a(C(x^a), p_i = 0)$ for $x^a = x^b$. This shows that

either a crossing of $U_a(C(x^a), p_i = 1)$ and $U_a(C(x^a), p_i = 0)$ exists for $x^a = x^b$, or $U_a(C(\overline{x}), p_i = 1) > U_a(C(\overline{x}), p_i = 0)$ which implies that $x^a = \overline{x}$ is an equilibrium. In each case, the high-motivation equilibrium is unique.

This proves that a high-motivation exists for all values of $w^A \in (\underline{x} - v(1), \overline{x})$. Moreover, in the high-motivation equilibrium, $\int_I p_i \to 0$ as $w^{A+} \to (\underline{x} - v(1))$ and $\int_I p_i \to 1$ as $w^{A-} \to \overline{x}$. And since both equilibrium cutoff values, x^b and x^a, are continuous in w^A, a high-motivation equilibrium with $\int_I p_i = \nu$ exists from some value of w^A.

Case (ii), low-motivation: Similar to Case (i), high-motivation, given $\nu < -(1 - \lambda)(\underline{x} + v(0))$, a market-clearing equilibrium exists where only non-motivated individuals select into institution A. For $w^A \in (\underline{x}, \underline{x} + v(0)$, an equilibrium exists where $x^a = \underline{x}$ since $w^A + v(0) < \underline{x}$. Moreover, over $w^A \in (\underline{x}, \underline{x} + v(0))$, $\int_I p_i$ is increasing continuously from 0 to $-(1 - \lambda)(\underline{x} + v(0))$ in this low-motivation equilibrium, which implies a market-clearing equilibrium exists since $\nu < -(1 - \lambda)(\underline{x} + v(0))$. ■

Proof of Lemma 4: First, we give an expression for C_t as a function of the cutoff types:

$$C_t = \frac{\int_I \hat{p}_{i,t} a_i}{\int_I \hat{p}_{i,t}} = \frac{\lambda(x_t^a - \underline{x})}{(1 - \lambda)(x_t^b - \underline{x}) + \lambda(x_t^a - \underline{x})}$$

Due to the quasi-linearity of both type's utility with respect to the wage, for interior values $\partial x_t^a(w_t^A, C_{t-1})/\partial w_t^A = \partial x_t^b(w_t^A)/\partial w_t^A = 1$, which implies that:

$$\partial C^t / \partial w_t^A = \frac{\lambda((1 - \lambda)(x_t^b(w_t^A) - \underline{x}) + \lambda(x_t^a(w_t^A, C_{t-1}) - \underline{x})) - \lambda(x_t^a(w_t^A) - \underline{x})}{((1 - \lambda)(x_t^b(w_t^A) - \underline{x}) + \lambda(x_t^a(w_t^A, C_{t-1}) - \underline{x}))^2}$$

This expression if negative iff:

$$(1 - \lambda)x_t^b(w_t^A) + \lambda x_t^a(w_t^A, C_{t-1}) < x_t^a(w_t^A),$$

which is true iff $x_t^a(w_t^A, C_{t-1}) > x_t^b(w_t^A)$.

Next, note that the relationship between $x_t^a(w_t^A, C_{t-1})$ and $x_t^b(w_t^A)$ depends only on the sign of $v(C_{t-1})$, since motivated types' utility is separable with regard to the wage and reputation. In particular:

$$x_t^a(w_t^A, C_{t-1}) \lesseqgtr x_t^b(w_t^A, C_{t-1}) \text{ iff } v(C_{t-1}) \lesseqgtr 0.$$

Lastly, note that the same relationship holds when one of the two cutoffs is non-interior, and when both are non-interior, $\partial C^t / \partial w_t^A = 0$. ■

Proof of Corollary 2: (1) follows directly from Lemma 4 since for $\{w_0^A, C_0\}$ to be dynamically stable, $v(C_0) \geq 0$, which implies $\partial C_t(w_t^A)/\partial w_t^A \leq 0$. This in turn implies $C_t \leq C_0$ for all for all t.

(2) Assume there exists $\{w_t^A\}$ such that $w_{t+1}^A \geq w_t^A$, and $C_t \geq C^*$ for some t. Take t equal to $\min_t\{t|C_t \geq C^*\}$. It follows that $C_{t-1} < C^*$, and therefore $\partial C_t(w_t^A)/\partial w_t^A > 0$. By Lemma 4, this implies that $v(C_{t-1}) \leq 0$. However, $v(C_{t-1}) \leq 0$ in turn implies that $C_t \leq \lambda < C^*$. ■

Proof of Proposition 2: Existence of a transition given $v(\lambda) > 0$ follows from the example provided in the main text.

Non-existence given $v(\lambda) = 0$ follows as a corollary to the proof of Lemma 4. By contradiction, assume $v(\lambda) = 0$, $v(C_0) < 0$ and $\{w_t^A\}$ such that a transition to $\{w_t^{A*}, C^*\}$ is an equilibrium. Since $v(C_0) < 0$, it follows that $C_1 < \lambda$. Therefore, for a transition to exist, it must be true that $C_t \leq \lambda$ and $C_{t+1} > \lambda$ for some t.

However, if $C_t = \lambda$, then $C_{t+1} = \lambda$ since $x_{t+1}^a(w_{t+1}^A, C_t) = x_{t+1}^b(w_{t+1}^A)$ when $v(C_t) = 0$. If $C_t > \lambda$, then $v(C_t) < 0$ and by the proof of Lemma 4 $C_{t+1} < \lambda$, which contradicts $C_{t+1} > \lambda$. ■

Proof of Proposition 3: For the proof of (i), note that for a transition to occur along $\{w^A\}$, it must be true that $C_{t-1} \leq \lambda$ and $C_t > \lambda$ for some t. Since $v(C_{t-1}) < 0$, it follows that $O_i^t(a_i = 1, x_i') > O_i^t(a_i = 0, x_i')$; that is, holding private-sector ability constant, the option value of public sector employment must be higher for a motivated worker than a non-motivated worker in period t. Additionally, for the option value of the motivated worker to be higher, it must be true that workers believe that the public sector will be over-demanded for some set $\{t+1, ..., t+n\}$ *and* $C_{t'} > \lambda$ for some $t' \in \{t+1, ..., t+n\}$.

However, given $\{w^A\}$, take the set of beliefs $\{\{\tilde{C}_{t+1}, ..., \tilde{C}_{t+n}\}\}$ such that $\tilde{C}_{t'} \leq \lambda$ for all $t' \in \{t+1, ..., t+n\}$. With any beliefs in this set, $O_i^{t'}(a_i = 1, x_i') < O_i^{t'}(a_i = 0, x_i')$, implying that $C_{t'} \leq \lambda$ for $t' \in \{t+1, ..., t+n\}$. Therefore, under these beliefs, $O_i^t(a_i = 1, x_i') < O_i^t(a_i = 0, x_i')$, implying that some beliefs $\{\tilde{C}_{t+1}, ..., \tilde{C}_{t+n}\}$ in this set are "self-fulling," in the sense that they constitute an equilibrium where $C_t \leq \lambda$. This shows that for any $\{w^A\}$ that admits a transition in equilibrium, there is an alternative set of beliefs such that there is no transition.

For the proof of (ii), we first show that there exists a wage path and a corresponding equilibrium that transitions to $\{w^{A*}, C^*\}$.

Take $\{w_t^A\}$ such that:

1. $w_t^A + v(C_0) = \overline{x}$ for $t < t'$, where $t' = \min\{t | v(C_{t'-1}) > 0\}$.
2. w_t^A for all $t \geq t'$ gives $\int_I \hat{p}_i = \nu$, given an equilibrium sequence $\{\tilde{q}_t', \tilde{C}_t'\}_{t'}^{\infty}$ such that $\tilde{q}_t' = 1$ for all $t \geq t'$.

Note that t' such that $v(C_{t'-1}) > 0$ exists, since A's reputation is strictly increasing in t for $t < t'$. Specifically, given $w_t^A = \overline{x} - v(C_0)$ for all t, $\hat{p}_{i,t} = 1$ for all workers independent of expectations of q_t, and a measure workers of size $\delta\nu$ will join institution A in each period from the set of workers with $p_{i,0} = 0$. Since the average motivation of these workers is weakly greater than λ, this implies that the sequence $\{C_t\}$ converges to λ as $t \to \infty$. Note that this sequence is unique (independent of expectations), which implies that it convergences monotonically given the initial non-motivated workforce. Since $v(\lambda) > 0$, there exists t' such that $v(C_{t'-1}) > 0$ Moreover, this t' is unique since the equilibrium sequence of $\{q_t, C_t\}_0^{t-1}$ is unique.

Next we show that (2) gives an equilibrium that transitions to $\{w^{A*}, C^*\}$. First, we state the analogous result to Lemma 4:

Corollary 3
If $v(C_{t-1}) \leq 0$ and $O_i^t = 0$, then $\partial C_t(w_t^A)/\partial w_t^A \geq 0$.
If $v(C_{t-1}) > 0$ and $O_i^t = 0$, then $\partial C_t(w_t^A)/\partial w_t^A \leq 0$.

This result implies that, if $\tilde{q}_t' = 1$, the same comparative statics between reputation and wage hold as the case of full replacement ($\delta = 1$).

Assume that $\{\tilde{q}_t, \tilde{C}_t\}_{t'}^{\infty} = \{\tilde{q}_t', \tilde{C}_t'\}_{t'}^{\infty}$. Under these expectations, $O_i^t = 0$ for all $t \geq t'$, so that Lemma 3 holds. Next, to show that this market-clearing equilibrium transitions to the high-motivation steady state, we show that, analogous to the example in the proof of Proposition 2, $w_t^A > w_{t+1}^A$ and $C_{t-1} < C_t$ for $t \geq t'$.

Note that $C_{t'-1} \leq \lambda$ since $C_{t'-2} < \lambda$ and $\int_I \hat{p}i, t' - 1 = 1$. However, by Lemma 2, $C_{t'} > \lambda$ since $v(C_{t'-1}) > 0$. Since $C_{t'-1} < C_{t'}$, $w_{t'+1}^A$ must be smaller than $w_{t'}^A$, otherwise $\int_I \hat{p}i, t' + 1$ would be greater than ν. By Corollary 3, $C_{t'} < C_{t'+1}$ (the relationship is strict since either x^a or x^b is interior). The same argument holds for all $t > t'$, implying that the sequence of $\{w_t^A, C_t\}$ converges to $\{w^{A*}, C^*\}$.

Lastly, we show that given the wage path detailed above, *any* equilibrium transitions to some $C^* + \epsilon$, where $\epsilon \geq 0$. Note that we have already shown that $\{w_t^A, C_{t-1}\}_0^{t'-1}$ is unique; however, multiple equilibrium may exist for $\{w_t^A, C_t\}_{t'}^{\infty}$. Take the set of reputations in the market-clearing, transition equilibrium detailed

above to be $\{w^A_{t''}, C''_{t-1}\}$; we will show, by contradiction, that for any other equilibrium $\{w^A_{t''}, C_{t-1}\}$, $C''_{t-1} \leq C_t$, which proves the result.

Assume there exists an equilibrium $\{w^A_{t''}, C_{t-1}\}$ where $C''_t > C_t$ for some $t \geq t'$. First, note that $\{w^A_{t''}, C''_{t-1}\}$ is unique given expectations that $\tilde{q}_t = 1$: if $q_t = 1$ in all periods, then workers will simply select employment that maximizes their period payoffs, which will result in the path of $\{w^A_t, C_{t-1}\}^{\infty}_{t'}$ of the market-clearing transition. Therefore, other equilibria will only occur under expectations that $\tilde{q}_t < 1$ for some set of periods. However, given $w^A_{t''}$, institution A will only be over-demanded, $\tilde{q}_t < 1$, if $C''_{t-1} < C_{t-1}$.

Take $\underline{t}$ equal to the minimum value of t where $\tilde{q}_t < 1$, and $\bar{t}$ equal to the minimum value of t where $C''_t > C_t$. It must be the case that $\bar{t} > \underline{t}$, since $C''_t = C_t$ for $t < \underline{t}$, and since $\tilde{q}_{\underline{t}} < 1$ only if $C''_t < C_{\underline{t}}$. Moreover, given $C''_{\underline{t}} < C_{\underline{t}}$, $C''_{\underline{t}+1} < C_{\underline{t}+1}$ since $x^{a''}_{\underline{t}+1} < x^a_{\underline{t}+t}$ and x^b_t is unchanged since, given $w^A_{t''}$ is decreasing, the option value of $p_i = 1$ is zero for the non-motivated cutoff type. This in turn implies, by induction, that $C''_t < C_t$ for all $t \geq \underline{t}$. Therefore, given $\{w^A_{t''}\}$, all equilibria correspond to a sequence of $\{C_{t-1}\}$ that are bounded below (weakly) by $\{C''_{t-1}\}$, which contradicts the existence of an equilibrium $\{w^A_{t''}, C_{t-1}\}$ where $C''_t > C_t$. ■