The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers
www.cesifo.org/wp

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

Justin Valasek

CESIFO WORKING PAPER NO. 5599
CATEGORY 1: PUBLIC FINANCE
NOVEMBER 2015

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5599*

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

Abstract

Recent research has highlighted social image and identity concerns as factors that influence economic decisions. Given that an individual's choice of employment may be important for their social image, we consider a model of worker sorting into the mission-oriented or private sector with motivated agents who also value the collective reputation of their place of employment. The initial insight of the analysis is that, from the institution's perspective, there may exist both a high-reputation, low-wage equilibrium and a low-reputation, high-wage equilibrium, which raises the question of how an institution can transition between equilibria. Our main contribution is to characterize a dynamic wage path that will transition from a low-reputation to a high-reputation steady state: Importantly, the effect of wages on motivation depend on the initial reputation - starting from low-reputation, higher wages crowd in motivation, while starting from high-reputation, higher wages crowd out motivation. Therefore, a non-monotonic wage path is required to achieve a transition to the low-wage, high-reputation equilibrium - an initial wage increase to crowd in motivated workers, followed by a wage decrease to crowd out non-motivated workers. These results provide a novel explanation for empirical findings in developing nations that - in direct contrast to evidence from developed nations - public sector workers are less prosocial and higher wages weakly increase motivation. Lastly, we discuss the implication of our results for policy measures aimed at reforming an institutional culture of corruption.

JEL-Codes: D230, D730, L320.

Keywords: motivated workers, institutional reform, public sector.

Justin Valasek
WZB Berlin / Germany
justin.valasek@wzb.eu

November 9, 2015
Thanks to Rachel Kranton, Steffen Huck, Paul Heidhues, Rajshri Jayaraman, Sebastian Kodrisch, Frank Heinemann, Tom Cunningham, Shanker Satyanath, Miriam Golden, Macartan Humphreys, Suresh Naidu, Stephan Meier, Lea Cassar, Gani Aldashev, Maitreesh Ghatak, Ferdinand von Siemens, Elena Paltseva and seminar participants at the Berlin Behavioral Economics Workshop, the MPI for Tax Law and Public Finance, Lund University, Queen's University Belfast, and Non-Profits, Governments and Organizations Workshop for valuable comments and suggestions. Also, I would like to thank Anders Bach-Mortensen for excellent research assistance.

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 of 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 mission-motivation, this wage differential is unsurprising: in an efficient equilibrium where motivated workers – workers who are intrinsically motivated when working for mission-oriented institutions – disproportionately select into the public sector, wages are lower as these workers are compensated by non-pecuniary payoffs of working for a well-regarded public administration with a mission of providing collective goods (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).

However, this relationship between wages, mission and motivation is not universal: recent empirical research has shown that prospective public-sector employees in developing nations are weakly less prosocial than their peers (Hanna and Wang (2013) and Banuri and Keefer (2014)), and that higher wages may increase, rather than crowd out, motivation (Dal Bó et al. (2013)). These findings suggest that institutional mission might not be sufficient to capture the full range of non-pecuniary elements valued by motivated individuals – instead, it may be necessary to take a broader perspective on institutional identity in this context. In a seminar treatise on bureaucrats, Wilson (1989) states: "There are three kinds of [nonmaterial] rewards: a sense of duty and purpose, the status that derives from individual recognition and personal power, and the associational benefits that come from being part of an organization (or a small group within that organization) that is highly regarded by its members or by society at large."

That is, in the example of Sweden, motivated workers may be attracted to join the bureaucracy precisely due its good 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.[2] In turn, institutional

[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] Other, equally relevant explanations, include: (1) 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

identity is a function of both the mission and the culture of the institution – while a motivated worker might be attracted a job in a well-regarded NGO, they might be negatively disposed towards working for a police force widely viewed as corrupt.

Here, we explore a model of labor-market sorting from the perspective of a mission-oriented institution (or firm), where motivated agents value a broader interpretation of institutional identity that also includes collective reputation – defined as average behavior within the institution (following Tirole (1996)).[3] We show how this model can account for contrasting empirical findings and derive implications for how an institution can reform, say, a culture of corruption using a commonly accessible policy tool – wage.

To give a simple illustration of the novel element of our model, framed in the context of corruption, we assume motivated workers derive positive value from a collective reputation for low corruption, and a negative value from a collective reputation for high corruption. While we remain agnostic as to the precise mechanism behind this behavioral element, the model we construct is consistent with a micro-foundation based on identity payoffs, prosocial signaling, or homophily. 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 mission-oriented institution; and (ii) the motivated type values the collective reputation of the institution.[4]

We first show that the model implies multiple equilibria – both low-motivation equilibria and high-motivation (efficient) equilibria may exist for given parameter values. This multiplicity is intuitive, given that motivated types effectively value an assortative labor-market match. Generally, a high-motivation equilibrium is characterized by a higher proportion of motivated types and a *lower* institutional wage than in the low-motivation equilibrium. The reason a lower wage is maintained in the high-motivation equilibrium is that motivated types are compensated by the collective reputation for high-motivation (in addition to any mission payoffs), while the low wage deters non-motivated types from entering the mission-oriented institution.[5]

However, in contrast to previous analyses, this does not imply that a high-motivation equilibrium can be achieved by simply setting a low wage. The low wage is a feature, rather than a cause, of the high-motivation equilibrium. Instead, the effect of a wage change

its employees' type; (2) Since the collective reputation and workforce composition of an institution are correlated, value homophily in the workplace (a la Lazarsfeld and Merton (1954)) provides yet another explanation why workers may value collective reputation.

[3] In our model, aggregate behavior is a linear function of the proportion of motivated individuals, which allows us to define collective reputation as either aggregate behavior or aggregate work-force composition.

[4] Non-motivated types may also value the collective reputation of the mission-oriented 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.

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

depends on the initial starting point: locally, in the low-motivation equilibrium a decrease in the wage *decreases* aggregate motivation, while in the high-motivation equilibrium a decrease in the wage *increases* motivation. The intuition behind this result lies in the fact that, holding aggregate motivation 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 low aggregate motivation, the proportion of non-motivated workers increases with a wage decrease, leading to an decrease in average motivation and making the mission-oriented institution even less attractive to motivated workers. By the same mechanism, however, starting from a case of high aggregate motivation, a lower wage increases average motivation. These findings organize the data that suggest a public-sector efficiency equilibrium in developing nations, but the reverse sorting and comparative statics in some developing-nations contexts where the public sector may not be highly regarded.

Several articles have highlighted examples of multiple equilibria in models with motivated agents (see for example Caselli and Morelli (2004), Macchiavello (2008), Kosfeld and von Siemens (2011), and Aldashev et al. (2015)) – the most novel contribution of our paper is that we formally analyze the problem of transitioning from a low-motivation equilibrium to a high-motivation equilibrium. We achieve this 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 institution in question. The policy tool we consider for enacting a transition is the relative wage; wages can be changed transparently and are a commonly-utilized policy tool for instituting public-sector reform (e.g. Besley and McLaren (1993); for a related paper that considers the use of imperfectly-enforceable laws to transition between steady-state equilibria, see Acemoglu and Jackson (2014)). Additionally, to ensure 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, we assume that motivated workers value the *lagged* collective reputation of the institution.[6]

We then characterize a wage path that induces a transition between a low-motivation equilibrium to a high-motivation equilibrium. We find that such a wage path generally involves an initial increase in the wage to attract more motivated types into the mission-oriented institution (crowding in motivation), followed by a gradual decrease of the wage to make the institution less attractive to non-motivated types (crowding out non-motivation).

[6]This reflects the notion that reputations are sticky – in a seminal article on collective reputation, Jean Tirole (1996) states "...stereotypes are long-lasting because new members of a group at least partially inherit the collective reputation of their elders." This assumption also serves as an approximation of a continuous-time model, where only atom-less groups of workers join the mission-oriented institution at any moment. Alternatively, an overlapping-generations approach could be used to similar effect, as in Acemoglu and Jackson (2014) and Acemoglu and Jackson (2015).

The intuition for a non-monotonic wage path follows from the comparative statics outlined above: Starting from a point of low motivation, only an increase in the wage will increase average motivation. Wage increases alone, however, cannot transition to the efficient, low-wage equilibrium. A transition can only be achieved if a tipping-point threshold of motivation 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 high-motivation, low-wage equilibrium.

We emphasize that the framework we analyze is not peculiar to the public sector and NGOs: 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 (for discussion of how identity and reputation affect employee selection in private-sector firms, see Henderson and Van den Steen (2015)); 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 à 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,[7] or that transitions are only possible for departments at universities with an overall reputation for academic excellence.

We consider several relevant extensions to the baseline model. First, we show how access to commitment may enable even a generic firm to transition to a high-motivation equilibrium, as long as workers are also optimistic about the future reputation of the institution. We then consider the case of correlation between a worker's ability and their level of motivation, and show how this correlation can be leveraged to achieve a transition. This case also provides insight into a commonly-attempted strategy of creating "elite" divisions within an existing institution – for this strategy to be successful, the institution must both recruit disproportionately from an ability type with a high average level of

[7]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."

motivation *and* offer a relatively high wage. Without an initial high wage, the strategy of recruiting from an ability type with a high average level of motivation may not be sufficient, since the overall reputation of the institution introduces an adverse selection problem. Lastly, we provide a discussion considering the application of our results to attempts to reform institutions with a reputation for corruption.

1.1 Literature

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), Prendergast (2008), Huck et al. (2012), and Fischer and Huddart (2008)). A subset of this literature considers motivation in the workplace, and has largely focused on optimal contracting in the presence of a motivated type, given that non-monetary incentives can be crowded out or distorted by traditional monetary incentive contracts (e.g. Murdock (2002), Dixit (2002), Sliwka (2007) and Ellingsen and Johannesson (2008); see Francois and Vlassopoulos (2008) and Prendergast (2008) for an overview).

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 mission-oriented sector, given the presence of different behavioral types. Several papers highlight that the efficiency wage in the mission-oriented 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, say, public-sector employment (Handy and Katz (1998), Francois (2000), and Besley and Ghatak (2005)). This result has been extended to account for the fact that other facets of the public sector may disproportionately attract individuals with harmful qualities, such as laziness or antisocial motives (Delfgaauw and Dur (2008), Auriol and Brilon (2014)), or into positions where altruism is counter-productive (Prendergast (2007)).

Additionally, several other papers have detailed mechanisms by which motivation can lead to multiple equilibria (see for example Caselli and Morelli (2004), Macchiavello (2008), Kosfeld and von Siemens (2011), and Aldashev et al. (2015)). In contrast to these papers, however, by considering motivated agents who value the collective reputation of an institution, 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 behavior, which implies that the impact of incentives becomes sensitive to the institution's starting point: higher wages increase motivation in

a low-motivation equilibrium, but decrease motivation in a high-motivation equilibrium. Therefore, reforming, say, a culture of corruption requires a more complex approach than simply replicating the incentives of a low-corruption institution.

Therefore, the main contribution of this paper is related to a set of papers that discuss mechanisms for transitioning between norms. In a seminal contribution on collective reputations, Tirole (1996) demonstrates how an extended anti-corruption campaign is required to shift to a dynamic path that leads from a high-corruption to a low-corruption steady state. The prediction of our model has a similar flavor: starting from a point of low motivation (analogous to high corruption), a wage increase must be sustained until a tipping-point reputation is reached, at which point the dynamics of the model lead to a path that converges to the high-motivation steady state. More recently, Besley et al. (2014) consider a dynamic model of tax compliance norms and demonstrate how, as new laws are introduced, endogenous social norms explain a lag in compliance. Bidner and Francois (2013) and Acemoglu and Jackson (2015) detail how transitions between norms of, respectively, political accountability and cooperation can occur. In contrast to previous papers, our contribution examines the role of selection in changing the culture of an institution, and offers policy prescriptions on how transitions can actively be enacted.

In a paper very related to ours, Acemoglu and Jackson (2014) detail how endogenously-enforced laws can be changed dynamically to transition between a steady-state of lawlessness to a steady-state of law-abiding. They show that a sudden shift in laws away from the current norm of behavior can be counter-productive, but that a series of incremental shifts can result in a transition. Our paper also shows that the path of a reform matters – a transition cannot be enacted by skipping straight to the wage of the efficient equilibrium. However, in contrast to the findings of Acemoglu and Jackson (2014), we find that the policy tool we consider (wages) must take a non-monotonic path for the system to transition to the optimal steady state.

Lastly, we argue that our results help reconcile the well-known policy prescription of a high public-sector "efficiency wage" to deter corruption (Besley and McLaren (1993)) with the concern that higher wages will crowd out intrinsic motivation (Besley and Ghatak (2005)). Indeed both results have empirical support: higher public-sector wages are weakly correlated with lower corruption (Treisman (2000), Van Rijckeghem and Weder (2001), and Di Tella and Schargrodsky (2003)), and there is evidence for a below-market "public-sector efficiency wage" (see Gregg et al. (2011)) in developed nations. However, in certain developing-nations contexts, recent empirical research has shown that prospective public-sector employees in developing nations are weakly less prosocial than their peers (Hanna and Wang (2013) and Banuri and Keefer (2014)), and that higher wages may increase, rather than crowd out, motivation (Dal Bó et al. (2013)). The theoretical results of this paper help reconcile these empirical findings: In contexts where public institutions have

a good reputation, higher wages will simply crowd out motivated workers. However, in contexts where public institutions have a poor reputation, say, due to high levels of corruption, then wage increases help combat corruption directly through the efficiency-wage argument, and indirectly by increasing the average motivation of public-sector workers.

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, which may be mission-oriented, 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 motivated 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 $y_i = 1$ (this assumption is relaxed in Section 5.2), 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, or net social output, 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. Institution 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.

We add one assumption about institution A's profit function, relative ν. Take x' that

solves:

$$(1-\lambda)\frac{x'-\underline{x}}{\overline{x}-\underline{x}}=\nu. \tag{1}$$

Assumption 1

The output institution A earns from each non-motivated worker, π, is greater than $\nu + x'$.

This assumption ensures that, holding average motivation constant, institution A always maximizes profit at full employment; that is, loosely, given employment less than ν, the marginal benefit of increasing w^A, and hence increasing employment, outweighs the cost of the increased wages.

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 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). Since the the collective reputation of B is perfectly negatively correlated with C, we can interpret C as A's reputation relative to B's (we do not explicitly consider payoffs associated with the collective reputation of B).

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 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.

To be consistent with the intuition that motivated workers place a positive value on a high reputation, and a negative value on a low reputation, I restrict the analysis to $v(1) > 0$ and $v(0) < 0$. Additionally, I highlight the analysis of the case where $v(\lambda) > 0$: Given a "generic" firm, it may be natural to assume that, holding constant wage and reputation, a motivated agent perceives employment in firm B and employment in firm A as equivalent $v(\lambda) = 0$. However, the main focus of the paper is on "mission-oriented" institutions, where motivated agents are directly motivated by the mission, or product, of institution A (as in Besley and Ghatak 2005). In our model, given the constant product produced by each worker in institution A, 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 be captured by the following assumption: $v(\lambda) > 0$, where $v(\lambda)$ represents the mission-benefits a motivated worker receives from employment in the mission-oriented institution.[8]

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 Nash. 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

[8] In Section 5.1, we characterize the case of a generic firm ($v(\lambda) = 0$) and show that mission-orientation is a necessary condition for reform.

$U_a(w^A, C, p_i = 1) \geq x_i$, and $C = \int_I \hat{p}_i a_i / \int_I \hat{p}_i$.[9]

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.

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 mission-oriented institution and institution B, 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 ($C(x^a)$).

First, we show that a high-motivation equilibrium exists for all w^A.

[9] A more precise definition of equilibrium would include the set of employment outcomes, $\{p_i\}$; however, to allow for more efficient notation, we use the fact that with a continuum of workers, C can be defined as a function of $\hat{p}_i$.

Lemma 2 (Existence high-motivation equilibrium)
Given $w^A \in (\underline{x}, \overline{x})$, there exists a unique high-motivation equilibrium, $C^h > \lambda$.

Since $v(C(x^a))$ is increasing and concave in x^a ($C(x^a)$ is concave), either an interior intersection exists, or a corner equilibrium exits ($\underline{x} > w^A + v(0)$ or $\overline{x} \leq w^A + v(1)$).

Next, we will partially characterize the optimal equilibrium from the point of view of institution A – similar to Besley and Ghatak (2005) we consider the objective of maximizing the efficiency of the mission-oriented sector. We will refer to an equilibrium as market-clearing if $\int_I \hat{p}_i = \nu$. However, even though a market-clearing wage is optimal given a fixed level of average motivation (by Assumption 1), as the following proposition illustrates, the equilibrium that maximizes $\pi^A(x^a, x^b, w^A)$ need not coincide with a market-clearing equilibrium.

Proposition 1 (Existence of optimal high-motivation equilibrium)
(i) A high-motivation, market-clearing equilibrium $\{w^{A^h}, C^{h'}\}$ exists.
(ii) The static equilibrium that maximizes $\pi^A(x^a, x^b, w^A)$, $\{w^{A^*}, C^*\}$, satisfies $w^{A^*} \leq w^{A^h}$, $C^* \geq C^{h'}$.

Existence of a high-motivation, market-clearing equilibrium follows from the fact that the cutoff abilities in the unique high-motivation equilibrium, $\{x^a, x^b\}^h$, are both continuous functions of w^A, which implies that a crossing with $\int_I \hat{p}_i = (1-\lambda)(x^b - \underline{x}) + \lambda(x^a - \underline{x}) = \nu$ exists.

For (ii), note that for any low-motivation equilibrium, there exists a high-motivation equilibrium with the same level of $\int_I \hat{p}_i$. Moreover, the w^A must be lower at this high-motivation equilibrium, due to the utility that motivated types receive from the higher level of reputation. Therefore, the equilibrium that maximizes social output, $\{w^{A^*}, C^*\}$, must be a high-motivation equilibrium. And since all high motivation equilibria with $w^A > w^{A'}$ have lower corresponding levels of reputation (see appendix for a formal proof), w^{A^*} must be smaller or equal to w^{A^h}.

Also note that since full employment is optimal for A given a fixed level of reputation, $w^{A^*} < w^{A^h}$ only if $\int_I \hat{p}_i^* a_i > \int_I \hat{p}_i^h a_i$. Formally, it is possible for $w^{A^*} < w^{A^h}$ since x^a can be a decreasing function of w^A – therefore, an equilibrium with a below-market wage may be optimal for A, since it may feature a greater absolute number of motivated workers.

Figure 1 illustrates equilibria for a given value of w^A; 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 $U_a(x^a, p_i = 0)$ and $U_a(x^a, p_i = 1)$ since at an intersection, given $C(x^a)$, motivated workers with $x_i = x^a$ are indifferent between employment in institutions A and B.

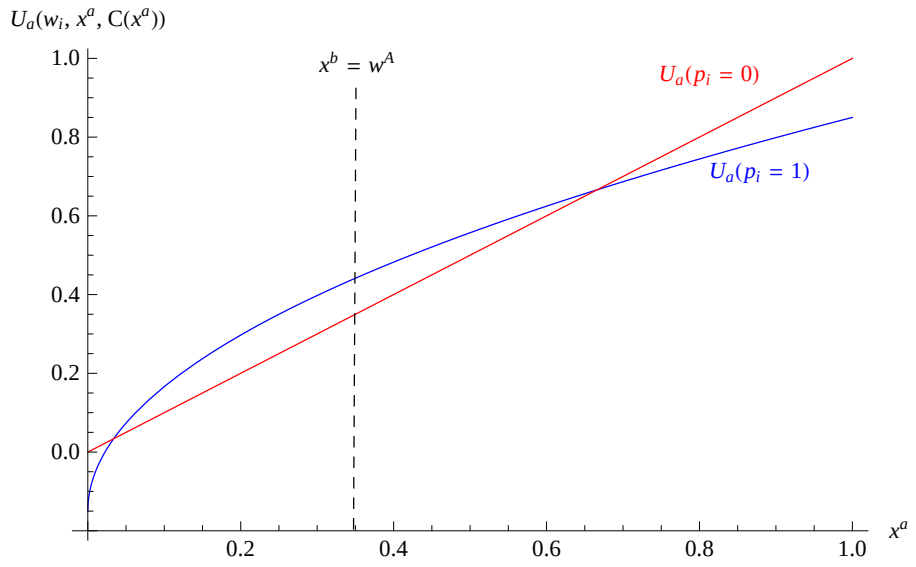

Figure 1: This graph illustrates the respective utility of employment in A and B for a motivated type with $x_i = x^a$, *given* that all motivated workers with $x_i < x^a$ set $\hat{p}_i = 1$. Therefore, given $x^b = w^A$, $C(x^a)$ is increasing with x^a.

Figure 1 also illustrates that multiple equilibria may exists for some values of w^A, and by Lemma 2, equilibria other than C^h must be low-motivation. The following proposition characterizes a sufficient condition for multiple equilibria.

Proposition 2 (Existence of multiple equilibria (high/low-motivation))
If $\nu < -(1-\lambda)(\underline{x} + v(0))$, *then:*
(i) A low-motivation market-clearing equilibrium $\{w^{A^l}, C^{l'}\}$ *exists, and* $w^{A^h} < w^{A^l}$.
(ii) A low-motivation equilibrium exists for all $w^A \leq w^{A^h}$.

Proposition 2 illustrates that multiple equilibria exist when institution A's demand for labor is small relative to the overall labor market, and when the motivated type places a high valuation on reputation ($v(0)$ is high).

The existence of multiple equilibria implies that the model is indeterminate on how the optimal equilibrium can be reached by the mission-oriented institution, since both high and low-motivation equilibria exist for w^{A^*} when ν is small. Moreover, it suggests that institutions that simply set wages at market-clearing levels could find themselves with either low or a high levels of motivation, depending on the underlying beliefs of agents as to which equilibrium the market will settle on (for a detailed discussion of the indeterminacy of policy analysis in the presence of multiple equilibria, see Morris and Shin (2000)).

Therefore, instead of relying on criteria for equilibrium selection, we instead consider the problem of transitioning from a point of low-motivation to the optimal high-motivation

equilibrium. Arguably, characterizing such a transition is a first-order concern, since the persistence of collective reputation (see Tirole (1996)) makes it unlikely that institutions in a high-motivation equilibrium will suddenly shift to a low-motivation equilibrium. However, the persistence of collective reputation, combined with legal and transactional constraints that prevent institutions from quickly replacing their entire workforce, also implies that moving from a low-motivation reputation to the optimal high-motivation equilibrium is likely to require a transition rather than a sudden shift. Therefore, in the next section, we take a similar approach to Tirole (1996) and Acemoglu and Jackson (2014) and introduce a dynamic version of the model that accounts for frictions stemming from repetitional-persistence and employment tenure, and that allows us to characterize a path for reforming a mission-oriented institution with a culture of low-motivation.

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. We also introduce two plausible sources of fiction to the dynamic model (a simple repetition of the static game would produce the same equilibria as in the static game).

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 et al. (2014), see Tirole (1996) for arguments as to why collective reputation cannot adjust instantaneously). 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. Formally, this assumption introduces the possibility that employment in institution A has an option value. 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, which we interpret as a natural rate of turnover due to, for example, retirement. 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 is determined 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.[10]

Institution A sets wages for all periods, $\{w_t^A\}_1^\infty$ at the beginning of period 1. This implies that A has access to commitment – we discuss the robustness of the main results to the weakening of this assumption below. The timing of the period game is as follows:

1. A sets $\{w_t^A\}_1^\infty$ (period 1 only).
2. $\{w_t^A\}_1^\infty$, C_{t-1} observed by workers.
3. Workers choose $\hat{p}_{i,t} \in \{0, 1\}$.
4. Period utility ($p_{i,t}$) realizes.
5. δ workers replaced at random.

[10] 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:

$$O_i^t = (1-q_{t+1})\left[u(w_{t+1}^A, C_t, a_i) - x_i + (1-\delta)O_i^{t+1}\right]$$

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 sum of the expected 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.

Since we are concerned with the reform of an existing institution, 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 $a_i = 0, x_i < x^b$.

Equilibrium

Since information is complete, the equilibrium concept we use is sub-game perfect Nash Equilibrium. Additionally, we follow the selection-criterium of Gul et al. (1986) and assume that agents do not condition their choices on the past actions of sets of agents of measure zero, which insures that unilateral deviations by a single worker do not affect the actions of the other workers. Given a set of wages, $\{w_t^A\}_1^\infty$, an equilibrium constitutes a set of employment choices, $\{\hat{p}_{i,t}\}_1^\infty$, that maximize each worker's dynamic utility:

$$U^t(w_t^A, \{\hat{p}_{i,t}\}, C_{t-1}, a_i, x_i, \{w_t^A\}, \{q_t\}, \{C_t\}),$$

given the implied reputation, $\{C_{t-1}\}_1^\infty$, and demand for jobs at the mission-oriented institution, $\{q_t\}_1^\infty$.

In equilibrium, expectations and outcomes must be consistent; however, for purposes of exposition it will occasionally be useful for us to explicitly refer to expectation over wages, reputation and demand for public sector jobs, denoted by $\{\tilde{w}_t^A\}$, $\{\tilde{C}_t\}$ and $\{\tilde{q}_t\}$.

4.1 Analysis of Dynamic Model

First, we look at an individual's decision rule, fixing $\{w_t^A\}$, $\{q_t\}$, and $\{C_t\}$. Each worker chooses $\hat{p}_{i,t} = 1$ if, and only if, the following expression holds:

$$u(w_t^A, C_{t-1}, a_i) + (1-\delta)O_i^t \geq x_i. \tag{2}$$

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

Again, given $\{w_t^A\}$, $\{q_t\}, \{C_t\}$, 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, by motivation-type, O_i^t is a monotonically decreasing function of x_i.

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

Lemma 3 (Motivated/Non-Motivated Reputation)
Given $q_{t+1} = 1$, 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 $q_{t+1} = 1$; 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 $w_t^A = \bar{w}^A$ for all t, an equilibrium $\{x_t^a, x_t^b, C_t\}_t^\infty$ is a steady-state equilibrium if $x_t^a = \bar{x}^a$, $\bar{x}_t^b = x^b$, and $C_t = \bar{C}$ for all t.

The relationship between static and dynamic equilibrium is clarified by the following Proposition:

Proposition 3 (Static Equilibrium $\Leftrightarrow$ Steady-State Equilibrium)
For each static equilibrium, there exists a corresponding steady-state equilibrium, and for each steady-state equilibrium, there exists a corresponding static equilibrium.

The proof of Proposition 3 follows trivially from the fact that, in a steady-state equilibrium, given a constant collective reputation, the option value of employment in A for the cutoff types must be equal to zero, which implies that $\bar{x}^a = \bar{w}^A + v(\bar{C})$, $\bar{x}^b = \bar{w}^A$. Note

that this implies that the conditions for a steady-state equilibrium in the dynamic setting are equivalent to the equilibrium conditions for a static equilibrium.

Proposition 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. Additionally, it gives the following corollary:

Corollary 1
The optimal stead-state equilibrium $\{w^{A*}, C^*\}_t^\infty$*, corresponds to the optimal static equilibrium,* $\{w^{A*}, C^*\}$*.*

In the following text, we use $\{w^{A*}, C^*\}$ to refer to the steady-state equilibrium. The following section analyzes the possibility of a dynamic transition from a low-reputation steady-state to $\{w^{A*}, C^*\}$, precipitated by a designer who sets wages in institution A, $\{w_t^A\}_1^\infty$.

4.2 Dynamic Transition

To formalize the problem of transition introduced at the end of Section 3, we address the question of whether a wage path $\{w_t^A\}_1^\infty$ exists that induces a transition in the state variable, C_t, from $C_0 < \lambda$ to $w_t^A = w^{A*}$ and $C_t = C^*$, where $\{w^{A*}, C^*\}$ is the steady-state equilibrium that corresponds to the static equilibrium that maximizes net social output, π^A.

Note that in this section, we do not explicitly seek the wage path that maximized the present value of net social output; however, if a transition to the optimal steady-state equilibrium is possible, then there exists a discount rate low enough such that a transition to $\{w^{A*}, C^*\}$ results in a higher present value of net social output relative to any other steady-state equilibrium (to be precise, any other steady-state equilibrium outside of an ϵ-neighborhood of $\{w^{A*}, C^*\}$), even if the transition is costly in the short-term. Additionally, we discuss optimal transition paths after characterizing conditions under which a transition exists.

Lastly, while the analysis is general, we often refer to the case where C_0 corresponds to a low-reputation market-clearing steady-state $(\bar{C}_0, \bar{w}_0^A)$. In these cases, we refer to $\bar{w}_0^A$ as the starting wage, and our description of wage-path includes $\bar{w}_0^A$.

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 main 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.

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 $v(C_{t-1}) > 0$.[11]

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

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

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

Here, we consider the case where the optimal steady-state equilibrium, $\{w^{A*}, C^*\}$, corresponds to a unique market-clearing steady-state equilibrium, $\{w^{A^h}, C^{h'}\}$. (Other cases, such as when $\{w^{A*}, C^*\}$ corresponds to $\int_I \hat{p}_{i,t} < \nu$, are addressed in the appendix.) Here, existence follows by construction, given the following example:

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.

[11] 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, the result holds more generally for changes in the wage that are large enough.

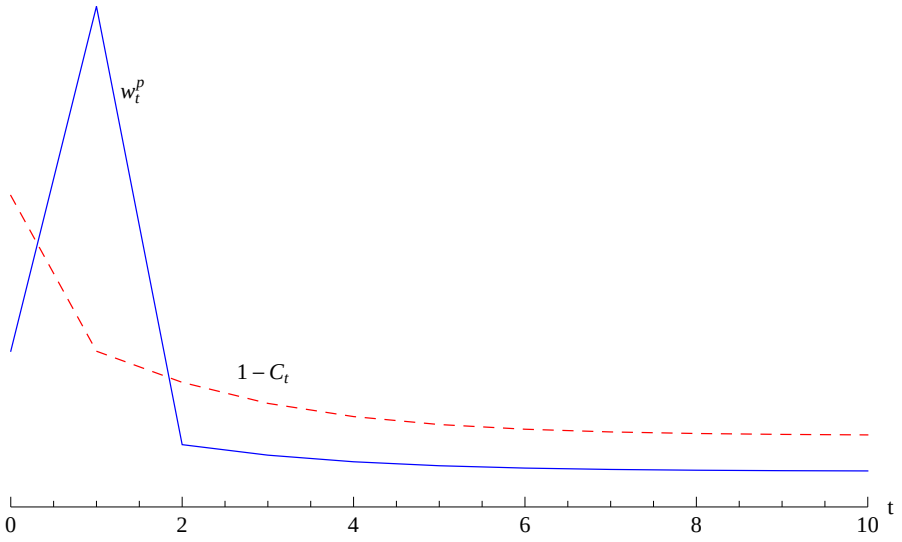

Figure 2: This graph illustrates a wage path that transitions from a high-corruption to a low-corruption equilibrium. Note the initial increase in the wage in the mission-oriented institution (w_t^A; solid line), followed by a decrease and convergence to the efficiency wage. Corruption ($(1 - C_t)$; dashed line), however, decreases monotonically.

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 employment 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 2). 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 close to 1 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.

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

The intuition from the example with $\delta = 1$ 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 5 (Existence of transition $v(C_0) < 0 \Rightarrow \{w^{A*}, C^*\}$)
If $v(C_0) < 0$, then there exists $\{w_t^A\}'$ such that $w_t^A \to w^{A*}$ and $C_t \to C^*$ (transition exists).

Proposition 5 shows that an equilibrium transition is always possible for a mission-oriented institution. First note that an equilibrium exists that transitions to $\{w^{A*}, C^*\}$ that is analogous to the example given for $\delta = 1$, for cases where $\{w^{A*}, C^*\}$ corresponds to a unique high-motivation market-clearing steady-state, $\{w^{A^h}, C^{h'}\}$. 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. This implies a stronger result than Proposition 5, namely that in all equilibria, $\{w_t^A\}'$ transitions to some $C' \geq C^*$.

Proposition 6 ($v(\lambda) > 0$: Transition in all Equilibria)
If $v(C_0) < 0$, then there exists $\{w_t^A\}'$ such that $w_t^A \to w^{A*}$ and $C_t = C^* + \epsilon$ for some t, $\epsilon \geq 0$, or $C_t \to C^* + \epsilon$ in any *equilibrium. (transition in any equilibrium).*

Optimal Transition Path

The precise tradeoff between speed of the transition and its cost depends on the relevant discount rate and other parameters. Moreover, many institutions may face legal and budgetary constraints that limit their ability to raise wages in any given period. Therefore, we detail the transition path that minimizes the maximum wage bill of institution A:

Proposition 7 (Minimum Budget)
The following wage path minimizes the maximum budget required to transition from any C_0 such that $v(C_{t-1}) \leq 0$ to $\{w^{A^*}, C^*\}$:

$$w_t^A = \begin{cases} \underline{x} + v(0) & \text{for } t \text{ s.t. } v(C_{t-1}) \leq 0 \\ w^{A'} \text{ where } w^{A'} \text{ solves } C_t = C^* & \text{for } t' = \min\{t | v(C_{t-1}) > 0\} \\ w^{A^*} & \text{for } t > t'. \end{cases}$$

The proof of proposition 7 can be demonstrated simply using the best-response dynamics of the static model since, given a fixed wage, an equilibrium of the dynamic model exists where the period-equilibrium converges to a stable equilibrium of the static model. Moreover, since only static equilibria that are corner solutions, or where $U_a(x^a, p_i = 1)$ crosses $U_a(x^a, p_i = 0)$ from above, are stable equilibria, given $v(\lambda) > 0$ (so that $U_a(x^a, p_i = 1) > U_a(x^a, p_i = 0)$ at $x^a = \lambda$) and $w_t^A = \underline{x} + v(0)$ (so that $x^a = \underline{x}$ is not an equilibrium), the unique stable static equilibrium is a high-motivation equilibrium. This implies that, given $w_t^A = \underline{x} + v(0)$, the dynamic model will converge to a high-motivation point.

In other words, if $w_t^A \geq \underline{x} + v(0)$, i.e. the wage is set high enough that the motivated worker with the lowest ability wishes to join the public sector, even when $C = 0$, then there is a unique stable equilibrium of the static model at $C' > \lambda$. This implies that, regardless of C_0, the dynamic model will converge to C', and the t' such that $v(C_{t'-1}) > 0$ will be reached in finite time.

Proposition 7 also shows that as soon as t' such that $v(C_{t-1}) > 0$ is reached, then a transition to the efficient steady-state can be achieved in a single period. The intuition is as follows: Given $v(C_{t-1}) > 0$, $x_t^a \geq x_t^b$, which implies that any $C_t \in [C_{t-1}, 1]$ can be achieved by setting w_t^A low enough. Put differently, by setting $w_t^A = \underline{x}$ (given an option value of 0), only motivated workers will remain in institution A in period t and $C_t = 1$, which implies that there exists $w_t^A > \underline{x}$ such that $C_t = C^*$. Of course, this transition path implies that institution A will not achieve a maximum profit in period t', since the institution is under-demanded. However, this may still be a profitable strategy, since it achieves a faster transition to the efficient steady-state equilibrium.[12]

[12]This strategy of transition may not be robust, since a large drop in salary may disturb the employer-employee relationship by, for example, erode trust. Therefore, we have highlighted transition paths that involve market-clearing wages.

Lastly, we point out that a transition can be achieved even if the mission-oriented institution is legally constrained to keep the wage of past employees constant or non-decreasing. As long as institution A can change w^A for incoming cohorts, then the wage paths described above will still result in a transition. In this case, however, a transition can only be achieved through the natural rate of replacement, δ, since non-motivated workers cannot be actively pushed out of their jobs through lower wages.

5 Extensions and Discussion

Here we discuss several relevant extensions and robustness checks.

5.1 Transition in Generic Firms ($v(\lambda) = 0$)

In this section, we show that $v(\lambda) > 0$ is not only a sufficient condition for a wage path that transitions to $\{w^{A*}, C^*\}$, it is also a necessary condition. Specifically, we consider the case of a "generic" firm without any mission-payoffs; i.e. where motivated workers are indifferent between institutions A and B when $w^A = x_i$ and $C = \lambda$, which translates to $v(\lambda) = 0$. (Results are analogous for $v(\lambda) < 0$.)

First, we state the analogous result to Proposition 2:

Proposition 2' (Existence of Market-Clearing Equilibria')
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.

Similar to Proposition 2, Proposition 2' shows that when $v(\lambda) = 0$, 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 ($v(1)$ is high).

Moving on to the dynamic model,

Proposition 4' (Existence of Transition)
Given $\delta = 1$, a wage path that transitions from $v(C_0) < 0$ ($C_0 < \lambda$) and $\{w^{A*}, C^*\}$ does not exist if $v(\lambda) = 0$.

The intuition for the nonexistence result for $v(\lambda) = 0$ follows from the same wage path illustrated in Section 4.1.1. Note that $C_1 = \lambda$ 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 $v(C_1) = v(\lambda) = 0$, and by Lemma 4 A's reputation is weakly increasing in w^A.

Proposition 4 and Proposition 4' 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, motivated workers prefer working in the institution given a neutral reputation. Therefore, $v(C_{t-1}) > 0$ can be achieved through a wage increase, which enables a transition that is unavailable to generic firms.

In the general case of $\delta \in (0, 1)$, a weaker result than Proposition 4' holds:

Proposition 5' (Existence of transition to $\{w^{A*}, C^*\}$)
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).*

That is, in contrast to the case of full replacement, Proposition 5' does not fully rule out the possibility of a transition when $v(\lambda) = 0$ – under certain conditions, 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 5.

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.

5.2 Exploiting Correlation between Ability and Motivation

It is natural to imagine that an agent's ability and their level of motivation may be correlated. For example, due to either selection or socialization, it is often suggested that individuals with a degree in economics are less prosocial and, depending on the mission of their workplace, may hence also be less motivated. Here, we consider the case where ability in institution A is heterogenous and correlated with motivation, and show that this extension of the model suggests potentially-important insights for transitioning to a high-motivation steady state.

We amend the baseline model by introducing heterogeneity in ability within institution A, y_i. Specifically, take $y_i \in \{y_1, ..., y_n\}$ and, for simplicity, there is a measure $1/n$ of each ability-type with compact index set I^n. Additionally, a proportion of λ^n of each ability-type is motivated (abusing notation, we use an n superscript to refer to variables that are differentiated by ability in institution A); take $\bar{\lambda}$ to be the average level of motivation of the population, $\bar{\lambda} = \sum^n \lambda^n$. To introduce correlation between ability and type, we assume that $\lambda^i \neq \lambda^j$ for some $i, j \in \{1, ..., n\}$.

We also allow for a correlation between y_i and x_i: given y_n, ability in institution B is uniformly distributed over $[\underline{x}^n, \overline{x}^n]$. Again, to focus on the problem of selecting based on motivation, ability (x_i, y_i) is observable and motivation is unobservable. Analogous to the model above, institution A has a unit demand of $\nu^n < 1/n$ of each ability-type, and sets a uniform wage conditional on y_i, $w^{A,n}$.

Lastly, take C_t^n to equal the average level of motivation by ability-level, and C_t^A equal to the average reputation of institution A:

$$C_t^n = \frac{\int_{I^n} a_i p_{i,t}}{\int_{I^n} p_{i,t}},$$
$$C_t^A = \frac{\sum^n \int_{I^n} a_i p_{i,t}}{\int_I p_{i,t}}.$$

The following proposition illustrates that, depending on the precise nature of reputation-payoffs, the correlation between ability and motivation can be exploited to transition between a low-motivation point and the high-motivation steady state equilibrium.

Proposition 8 (Transition of Average Reputation)
(i) If motivated agents value ability-contingent reputation, i.e. $U_{a,t}(w_i, C) = w_{i,t} + v(C_{t-1}^n)p_{i,t}$, *then a wage path that transitions from* $v(C_0) < 0$ *to* $\{w^{A*}, C^*\}$ *in all equilibria exists if, and only if,* $v(\lambda^n) > 0$.
(ii) If motivated agents value the average reputation of the institution, i.e. $U_{a,t}(w_i, C) =$

$w_{i,t} + v(C^A_{t-1})p_{i,t}$, *then a wage path that transitions from* $v(C_0) < 0$ *to* $\{w^{A*}, C^*\}$ *in all equilibria exists if* $v(\bar{\lambda}) \geq 0$.

Result (i) is a straightforward corollary of Proposition 5: if motivated agents value ability-contingent reputation, then each ability category can be treated as its own institution.

Result (ii), however, illustrates that if motivated agents value average reputation, then institution A can exploit the correlation between ability and motivation to transition even if $v(\bar{\lambda}) = 0$. This result follows from the simple intuition that A can manipulate its reputation by disproportionately hiring agents from ability levels with average levels of motivation above that of the population average.

Formally, a transition can be achieved with a wage path, $\{w^{A,n}_t\}$, where:

$$w^{A,n}_t \begin{cases} = 0 & \text{if } \lambda_n \leq \bar{\lambda} \\ = \overline{x}^n + v(C_{t-1}) & \text{if } \lambda_n > \bar{\lambda}, \end{cases}$$

until t' such that $v(C_{t'-1}) > 0$, and $w^{A,n}_t$ is market-clearing for $t \geq t'$. Note that, in contrast to the main analysis with $v(\lambda) = 0$, t' exists since an ability level exists with $\lambda^n > \bar{\lambda}$ by the assumption that $\lambda^i \neq \lambda^j$ for some $i, j \in \{1, ..., n\}$, and since $C_t \to \sum^{n'} \lambda^n / |n'| > \bar{\lambda}$ as $t \to \infty$, where $n' \equiv \{n | \lambda_n > \bar{\lambda}\}$. That is, a transition is enabled by the fact that institution can achieve a motivated reputation by selectively raising wages in ability-level that have a higher proportion of motivated types.

5.3 Reforming an Institutional Culture of Corruption

In this section, we discuss the implications of the formal results for public-sector reform. We emphasize the frame of corruption due to its saliency with regard to an institution's reputation – perceptions of corruption correlate with actual behavior (see Lambsdorff (2007)) – and for its social costs (Svensson (2005)). Specifically, we have in mind "petty corruption" within a public institution, such as a bureaucracy that is responsible for an allocation choice (allocating hospital beds based on need) or a monitoring task (firm/citizen compliance with laws, such as a health inspector monitoring restaurant compliance with health regulations, or police monitoring compliance with traffic laws). The objective function of the public institution is not to maximize the monetary benefit of the transacting parties, namely the public agent and individual citizen. Instead, the public institution hires the public agent to make a decision that has diffuse social benefits if the objective function is followed, such as higher levels of aggregate public health or safe streets, but implies targeted monetary costs, such as low-need individuals paying for own medical treatment or speeding motorists paying fines. Moreover, the matching of the decision to

the social objective function is imperfectly monitored.

The combination of imperfect monitoring and targeted monetary costs open up the possibility of a transaction between the public agent and the citizen, where a monetary transfer is made from the citizen to the agent in exchange for the agent *not* matching the decision to the underlying social objective. Commonly, this is referred to as corruption. In this setting, corruption is socially harmful because it inhibits the government from being able to implement a non-monetary social objective. Moreover, due to the diffuse benefits of following the social objective function, corruption cannot be prevented (efficiency cannot be achieved) by simply making the public agent a residual claimant.

Since individuals vary in the rate that they trade-off between social and personal benefits (Andreoni and Miller (2002)), a natural implication is that some individuals are less prone to corruption when employed by the public institution (this maps into the definition of motivation used above). Therefore, one for reforming an institutional culture of corruption is to attract the motivated individuals into the public sector. And, while the matching problem is only one dimension of an effective anti-corruption strategy, both theoretical and empirical research suggests that it is a potentially important dimension: Gregg et al. (2011) show that 'labor donations' (unpaid overtime) are higher in the non-profit sector. Moreover, individuals who switch from the non-profit sector to the for profit sector continue to donate at higher levels, suggesting that the difference is due to selection rather than socialization effects in the non-profit sector.[13]

When considering the example of corruption, an important caveat to this paper's prediction that higher wages should decrease corruption in high-corruption institutions is that the model assumes that, when the mission-oriented institution is over-demanded, jobs are randomly allocated to the applicants. However, it is possible that the process of allocating over-demanded public sector jobs is itself partially corrupt. In this case, the probability of receiving a public sector job might be an increasing function of the relative surplus that a worker receives from public employment, since some of this surplus must be shared with the employer. This would result in stricter conditions for the existence of a transition path since, starting from a point of high corruption (low-motivation), the utility surplus of a job in the mission oriented institution is relatively higher for a non-motivated worker than for the comparable motivated worker. This implies that for the initial wage increases to have a positive effect on the collective reputation of the mission-oriented institution, which is a necessary condition for a successful transition, it requires a minimal level of institutional capacity to ensure that some proportion of over-demanded jobs are allocated through a non-corrupt process.

Returning to Sweden as an example of a country that may be in a public-sector

[13]See Dur and Zoutenbier (2013) and Banuri and Keefer (2013) for additional evidence that motivated motivations differ between the public and private sector.

efficiency equilibrium, it is not the case that Sweden's bureaucracy has always been held in high esteem. In the mid-1800's corruption was endemic to the public administration, where the system of payments to government officials involved "gifts" for services rendered (see Rothstein (2011), 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.

While it is impossible to consider this a test of the model's predictions, it is still informative to consider whether there is any evidence that the Swedish transition followed a non-monotonic wage path to arrive at the current point of low relative wages and high efficiency. Interestingly, while public-sector wages were initially increased to ensure that bureaucrats were not dependent on direct payments from citizens, and to open the profession to a larger class of individuals, public-sector wages fell relative to private-sector wages in "the decades surrounding the Second World War" (Granholm (2013) pp. 101, translated from Swedish). However, Sundell (2014) provides evidence that nepotism in the Swedish bureaucracy decreased continually through this period, and argues that this is indicative of a continuous transition to a well-functioning public sector.

Lastly, while the example of Sweden is illustrative, it does not directly address the most novel prediction of the model – that the direction of the effect of wages on motivation is a function of the institution's collective reputation. To assess whether this prediction is consistent with empirical findings, we consider two recent randomized control trials (RTCs) that vary wages for mission-oriented institutions and measure the resulting effect on the motivation of applicants. Dal Bo et al. randomize the wages for a position as a Community Development Agent for a program in Mexico, a country that ranks in the bottom half of Transparency International's Corruption Perceptions Index, which is designed to capture perceptions of corruption in the public sector – the authors find that higher wages have a *positive* impact on public sector motivation. Deserranno randomizes wages for a position as a Community Health Promoter for a program in Uganda, a country that ranks well below Mexico in the Corruption Perceptions Index – the author finds that higher wages have a *negative* impact on prosocial motivation. A key difference between the two RTC's, however, is the nature of the employers: while the program in Mexico was directly administered by the Mexican government, and hence could be considered a public-sector position, the program in Uganda was administered by BRAC, which is second on The Global Journal's rankings of the top 500 NGOs (2015). Therefore, these findings are consistent with the predictions of the model, assuming that the Mexican government has a collective reputation for low motivation, while BRAC, as a highly respected NGO, has a collective reputation for high motivation.

6 Conclusion

We conclude with some brief comments on how 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 (see Bernheim (1994) and Huck et al. (2012) for models of social norms based on conformity), 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

Acemoglu, D and M O Jackson (2015), "History, Expectations, and Leadership in the Evolution of Social Norms." *The Review of Economic Studies*, 82, 423–456.

Acemoglu, Daron and Matthew Jackson (2014), "Social Norms and the Enforcement of Laws." Technical report, National Bureau of Economic Research, Cambridge, MA, Cambridge, MA.

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.

Aldashev, G, E Jaimovich, and T Verdier (2015), "When warm glow burns: Motivational (mis) allocation in the non-profit sector." *Working Paper*.

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 non-profit 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.

Besley, Timothy and John McLaren (1993), "Taxes and Bribery: The Role of Wage Incentives." *The Economic Journal*, 103, 119–141.

Bidner, Chris and Patrick Francois (2013), "The Emergence of Political Accountability*." *The Quarterly Journal of Economics*.

Caselli, Francesco and Massimo Morelli (2004), "Bad politicians." *Journal of Public Economics*, 88, 759–782.

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 ....*

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.

Granholm, Arne (2013), "När Professionerna Blev Byråkrater: Om statlig byråkrati under 200 år och hur traditionella professioners särställning minskat." *Umeå Papers in Economic History*.

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.

Gul, Faruk, Hugo Sonnenschein, and Robert Wilson (1986), "Foundations of dynamic monopoly and the coase conjecture." *Journal of Economic Theory*, 39, 155–190.

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.

Hanna, Rema and S Y Wang (2013), "Dishonesty and Selection into Public Service." *nber.org*.

Henderson, Rebecca and Eric Van den Steen (2015), "Why Do Firms Have "Purpose"? The Firm's Role as a Carrier of Identity and Reputation." *American Economic Review*, 105, 326–330.

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

Kosfeld, Michael and Ferdinand A von Siemens (2011), "Competition, cooperation, and corporate culture." *The RAND Journal of Economics*, 42, 23–43.

Macchiavello, R (2008), "Public sector motivation and development failures." *Journal of Development Economics*.

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

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.

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.

Wilson, James Q. (1989), *Bureaucracy: What Government Agencies Do and Why They Do It*. Basic Books.

7 Appendix A: Proofs

Proof of Proposition 1:
(i) First, 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.

(ii) Here we use the comparative statics of the unique high-motivation equilibrium with respect to w^A. Note that result follows if $\partial C(x^a, x^b)/\partial w^A \leq 0$, since the market-clearing high-motivation equilibrium will have higher net social output than an equilibrium with a higher wage and lower average motivation.

Starting from w^{A^h}, consider a wage increase of δw to $w^{A''}$. Note that $\delta x^b = \delta w$. Assume, by contradiction, that in the unique high-motivation equilibrium at $w^{A''}$, $C(x^{a''}, x^{b''})$ is equal to $C^{h'}$. This implies that $\delta x^a = \delta x^b = \delta w$, since the reputation-payoffs are the same after the wage increase, and utility is linear with respect to wage. However, if $\delta x^a = \delta x^b$, then $C(x^{a''}, x^{b''}) < C^{h'}$, since $C^{h'} > \lambda$. Therefore, at $x^{a''} = x^{a'} + \delta w$, $x^{a''} > w^{A''} + v(C(x^{a''}, x^{b''}))$, which implies that, given $w^{A''}$, the corresponding high-motivation equilibrium has $C'' < C^{h'}$. ■

Proof of Proposition 2:
(i) 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 $\nu < -(1-\lambda)(\underline{x} + v(0))$.

(ii) Follows trivially from the proof of (i) and from the fact that the wage in any low-reputation market-clearing equilibrium must be higher than w^{A^h}. ■

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 4:
Given $\{w^{A^*}, C^*\} = \{w^{A^h}, C^{h'}\}$ and $\{w^{A^h}, C^{h'}\}$ unique, existence of a transition given $v(\lambda) > 0$ follows from the example provided in the main text.

If this is not the case, then the following wage path transitions to $\{w^{A^*}, C^*\}$:

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_2^A solves $C_2 = C^*$.
3. w_t^A for $t > 2$ equals w^{A^*}.

Note that there exists w_2^A such that $C_2 = C^*$, since given $v(C_1 = \lambda) > 0$, $C_2(w_2^A)$ is a continuous function with a range of $[\lambda, 1]$. ■

Proof of Proposition 5:
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 4, $w_t^A > w_{t+1}^A$ 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 3,

$C_{t'} > \lambda$ since $v(C_{t'-1}) > 0$. Since $C_{t'-1} < C_{t'}$, $w^A_{t'+1}$ must be smaller than $w^A_{t'}$, 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 $\{w^A_t, C_t\}$ converges to $\{w^{A*}, C^*\}$. ■

Proof of Proposition 6:
Here we show that given the wage path detailed in the proof of Proposition 5, *any* equilibrium transitions to some $C^* + \epsilon$, where $\epsilon \geq 0$. Note that we have already shown that $\{w^A_t, C_{t-1}\}_0^{t'-1}$ is unique; however, multiple equilibrium may exist for $\{w^A_t, C_t\}_{t'}^{\infty}$. 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 $\{w^A_t, C_{t-1}\}_{t'}^{\infty}$ 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 $C''_t > C_t$. ■

Proof of Proposition 2':
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 $\nu < \lambda(\underline{x} + v(1))$.

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 $w^A = \underline{x} + \nu$. ■

Proof of Proposition 4':
Non-existence given $v(\lambda) = 0$ follows as a corollary to the proof of Lemma 4. By contra-

diction, 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 5':
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. ■