CIRJE-F-169

The Erosion and Sustainability of Norms and Morale

Michihiro Kandori

University of Tokyo

September 2002

The Erosion and Sustainability of Norms and Morale *

KANDORI, Michihiro†
Faculty of Economics, University of Tokyo

September 19, 2002

Abstract

The initially high performance of a socioeconomic organization is quite often subject to gradual erosion over time. We present a simple model which captures such a phenomenon. We assume that players are partly motivated by certain psychological factors, norms and morale, and they are willing to exert extra effort if others do so. This results in a "continuum" of equilibrium effort levels, whose minimum corresponds to the Nash equilibrium with respect to the material incentives. We show that repeated random shocks induce the erosion of equilibrium effort levels, but they do not completely decay; in the long run a certain range of efforts are sustainable. Our model shows that different organizations typically enjoy diverse norms and morale, which persist for a long time, in the vicinity of the equilibrium determined by material incentives.

JEL Classification Numbers: A12, A13, C70, C72, C73, C91, C92, D63, D64, H41.

*This is the text for the 2002 JEA-Nakahara Prize Lecture, delivered at the Fall meeting of the Japanese Economic Association at the University of Hiroshima on 13-14 October 2002.

†E-mail: kandori@e.u-tokyo.ac.jp. Tel: +81 3 5841 5657. Fax: +81 3 5841 5521. The author is grateful to J. Hofbauer, W. Sandholm, H. Imai, T. Kikutani, A. Okada, T. Sekiguchi, T. Shichijo, H. Suehiro, the seminar participants at the University of Kyoto, D. Shimizu and an anonymous referee for helpful comments and discussion. Financial supports from CIRJE at the University of Tokyo and the Japan Economic Research Foundation are also gratefully acknowledged.

1 Introduction

The initially high performance of a socioeconomic organization is quite often subject to gradual erosion over time. The waiting time to obtain referee's reports for professional journals has been ever increasing (at least in economics[1]), and a class tends to start later as a semester progresses[2]. In the experimental studies of the voluntary contributions to public goods, it has been repeatedly observed that contributions gradually decay over time. In the present paper, we construct a simple model which captures such a phenomenon. Our theory attributes the dynamics to certain psychological factors, which might be phrased as norms and morale. An organization may initially enjoy good performance due to high morale and effective work norms. Those psychological factors have reciprocal nature in the sense that one is able to maintain high morale and observe norms if others do so. However, one's behavior is usually subject to random shocks, and as a result the initial morale and norms are upset in due course. We show that a gradual erosion of morale and norms results through the interplay of material (pecuniary) incentives and the effectiveness of the psychological factors, under perpetual random shocks. An interesting question is whether morale and norms completely decay so that only material incentives matter in the long run. Our model predicts that this is not the case, and it shows that a certain *range* of morale and norms are sustainable in the long run. Hence our model shows that organizations enjoy diverse norms and morale, which persist for a long time, in the vicinity of the equilibrium determined by material incentives. Our model thus sheds light on the effectiveness, limitations, and diversity of norms and morale in resource allocation problems.

There could be an alternative explanation for the decay of performance in an organization, which is based on learning. The learning explanation maintains that agents are not fully aware of material incentives in the short run, but they gradually learn to behave as *homo economicus*: In the long run, they play the equilibrium determined by the material incentives. While learning is undeniably an important element in the dynamics of performance, there are some evidences contradicting such an explanation. In the public goods experiments where no contribution is the dominant strategy, a stylized fact is that contributions decay over time but non-negligible contributions remain even in the long run (Dawes and Thaler [6], Isaac, McCue, and Plott [13] and Isaac, Walker, and Williams [14]). A particularly revealing result is

[1] Ellison [8] and [9]. See Section 5 and footnote 16 for more detailed discussion.

[2] At least in my experience.

reported by Andreoni [4], who let subjects play the voluntary contributions game ten rounds. The average contribution declined from 19.9 to 5.3 (zero contribution is dominant). After the tenth round, the subjects are unexpectedly told to repeat the same experiment again. In the first round of the new experiment, the average shot back to 19.7. This indicates that the subjects are aware of material incentives, but there are non-material (i.e., psychological) factors which determine their behavior[3].

What could be the nature of the relevant psychological factors? In the growing literature on psychology and economics (or behavioral economics), basically three different formulations have been proposed. One is called *pure altruism*, which specifies that one's utility is a weighed sum of her own and others' payoffs. Another formulation is referred to as *warm glow*, which assumes that cooperative behavior *per se* provides positive utility. The third formulation is what Rabin [19] called *reciprocal altruism*, which is based on the idea that one is inclined to be nice to those who are nice to himself[4]. The decay of morale and norms are most naturally captured by this formulation, as we will see. Rabin [18] demonstrated a way to superimpose reciprocal altruism to material payoffs for 2×2 games, building on the notion of psychological games proposed by Geanakoplos, Pearce, and Stacchetti [12]. Levine [16] provided an alternative formulation, where attitude towards other players is treated as a given but privately known parameter. Lindbeck, Nyberg, and Weibull [17] presented a model of social welfare benefits in which the embarrassment to live off a welfare benefit is a decreasing function of the number of people on the benefit. Reciprocal altruism often produces multiple equilibra (mutually nice and mutually hostile ones, for example), but none of those papers address the dynamics of psychological factors to examine the relative stability and sustainability of multiple psychological equilibria. The purpose of the present paper is to complement those works with an explicit model of dynamics.

Let us now sketch the structure of our model. We look at a "social dilemma" situation, where the dominant strategy is Pareto inefficient. The

[3]There are some differences between our model and the public goods experiments. In the majority of experiments, material payoffs are linear in contributions, and individual contributions are not disclosed to the subjects. Those conditions are not met in our model. Also our model does not formally show how the initial norm of a (new) experiment is determined. Our purpose here is to present a simple model to capture the erosion of performance, and we believe that the basic logic and technique in this paper provide some insights into the experimental results. Constructing a model that closely reproduces the experimental results is an important future research adgenda.

[4]See also Fehr and Schmidt [10] for an important alternative formulation where one cares his own payoff relative to others.

strategy of a player is interpreted as his effort level, and it takes on a number of values. In addition to the material payoffs, we introduce psychological payoffs, parametrized by two factors, a norm and its binding power. A norm is the effort level that people expect themselves to exert, and a player suffers from a negative payoff if his effort falls short of the norm. The magnitude of the negative payoff, which we call the binding power of the norm, is the greater, the closer they follow the norm. Once such effects are introduced, the prisoner's dilemma like situation turns into a coordination game, as Rabin [18][19] stresses, and this is the first essential ingredient of our model. Depending on the relative strength of the material and psychological payoffs, the maximum equilibrium effort level is determined, and there are "continuum" of equilibrium effort levels, whose minimum corresponds to the Nash equilibrium with respect to the material payoffs. This is the second essential ingredient.

Now observe that each equilibrium is strict in the sense that a unilateral deviation strictly reduces one's payoff. Hence traditional equilibrium refinements concepts are ineffective to discriminate them, but the stochastic evolutionary game models proposed by Kandori, Mailath, and Rob [15] and Young [24] can be fruitfully applied to address the dynamic stability of equilibria. We will show that, due to random shocks, the system moves from equilibria with higher efforts to the ones with lower efforts. This is the third ingredient of our approach. Ellison [7] noted that the stochastic evolutionary models are particularly relevant in the case where the long run stochastically stable outcome is achieved via a series of small steps between intermediate steady states, as the waiting time to see the stochastic evolution effects can be realistically short. The kind of psychological factors we consider exactly produce such a game, a coordination game where the stable outcomes are achieved through "step-by-step" evolution over a "continuum" of equilibria.

2 The Social Dilemma with Norms and Morale

We consider an N-player version of the prisoner's dilemma game affected by psychological factors. We assume that player i's payoff is given by

$$u_i(\mathbf{e}, k, m) = \sum_{j=1}^{N} e_j - c(e_i) - k[m - e_i]_+, \tag{1}$$

where $e_i \in \{0, 1, 2, ..., L\}$ represents player i's effort level, $\mathbf{e} = (e_1, ..., e_N)$ is an effort profile, c is the cost of effort, and $[x]_+$denotes $\max\{x, 0\}$. As the

effort level is discrete, define (downward) marginal cost of effort level for $e = 1, 2, ..., L$ by

$$\Delta c(e) \equiv c(e) - c(e-1),$$

and assume that it is positive and strictly increasing. This is the marginal *benefit* of *reducing* effort, and it plays a crucial role in what follows. The first two terms of the right hand side of (1) captures the *material (or pecuniary) payoff*, and the last term represents the *psychological payoff*. One of the elements that determine the latter is m, which represents a *norm* among the workers. It can be thought as the acceptable level of effort, or the effort level that the players think they should exert. The player suffers from negative psychological cost if his effort level falls short of m. Parameter k denotes the strength of the psychological cost (or the *binding power* of the norm). We assume that in the equilibrium or the steady state,

$$m = \text{median}\{e_1, ..., e_N\} \text{ and}$$

$$k = K\left(\sum_{i=1}^{N}[m - e_i]_+\right),$$

where $K(\cdot)$ is a non-negative decreasing function[5]. For simplicity, we assume that the number of players is odd, so that there is a player whose effort level coincides with the median. One may interpret that (m, k) represents the *morale* of the players, where a norm m close to the efficient action and a strong binding power k correspond to high morale[6].

A couple of comments are in order about the above specification. We will consider the dynamic process where the norm m evolves over time, according to the actual effort levels taken by the players. A simplest formulation is that the norm at time t is determined by the effort levels at $t-1$. We may possibly use the average of the effort levels, but this suffers from some drawbacks. For example, if the effort profile $(e_1, ..., e_N)$ changes from $(10, 10, ..., 10)$ to $(3, 10, 10, ..., 10)$, the norm would immediately fall from 10 if the norm were defined as the average. It would be more natural, however, to suppose that the norm remains to be 10 and the

[5] In the present formulation, the binding power of the norm is affected only when players' efforts fall short of the norm. One may also assume that the binding power increases when some of the players exert higher levels of effort than the norm. Our analysis below is unaffected by such reformulation.

[6] The players may well enjoy satisfaction of high morale *per se*. To capture this effect, we may add to each player's utility a term $h(m, k)$, which is an increasing function of k and maximized, for any given level of k, when m is equal to the efficient effort level. Our analysis is unaffected by such reformulation.

deviation by the first player reduces the credibility of the norm. Our formulation captures such a mechanism; under our formulation, the norm remains $m = 10$ but the binding power of the norm k decreases. Also consider the case $e = (0, 0, 2, 7, 7, 8, 8, 8, 9, 9, 10)$. It would be more natural to recognize the cluster between 7 and 10 and expect the norm to be somewhere in the cluster. Our use of median is in line with this observation, deriving $m = 8$. In contrast, the average fails to recognize the cluster and provides $m = 6.1$. By definition, median always chooses an effort level in a cluster if a majority of the players are in the cluster[7]. Note that the average and the median minimize[8] $\sum_{i=1}^{N}(x - e_i)^2$ and $\sum_{i=1}^{N}|x - e_i|$ respectively, so that the latter places less weights on "outliers". To see another property of median, consider a change from $(1, 5, 5, 5, 10)$ to $(1, 5, 2, 5, 10)$. The median remains unchanged. Unlike the average, the median is generally *insensitive to any single player's effort, if there is a tight cluster of effort levels followed by a majority of players.* Hence it captures the *inertia* of the norm in a simplest possible way. Finally, the median is always in the strategy space $\{0, 1, ..., L\}$ (under our assumption of odd number of players) and make our analysis transparent.

Definition 1 *Effort profile* $\mathbf{e}^*$ *is a* morale equilibrium *if*

$$\forall i \ \forall e_i \ u_i(\mathbf{e}^*, k, m) \geq u_i(\mathbf{e}^*_{-i}, e_i, k, m),$$

$$m = median\{e_1^*, ..., e_N^*\} \text{ and}$$

$$k = K\left(\sum_{i=1}^{N}[m - e_i^*]_+\right).$$

This is somewhat different from the standard definition of Nash equilibrium of the game where i's payoff is given by $U_i(\mathbf{e}) = u_i(\mathbf{e}, k(\mathbf{e}), m(\mathbf{e}))$. In our definition, each player takes k and m given when assessing the gain from deviation. The parameters m and k, the norm and its binding power, are psychological factors reflecting mutual expectations of players, which are given at each moment (see the dynamics below for further motivation).

[7] This is true whatever the definition of cluster is, as long as it is a connected set of effort levels (i.e., as long as the cluster consists of all effort levels between e' and e", for some $e' < e$".)

[8] This is seen as follows. Let $x \neq e_1, ..., e_N$ and let n be the number of players whose efforts are below x. Denote the sum of the absolute value of the errors by E. Then we have $dE/dx = n - (N - n)$. Hence E is decreasing until x hits the median and then it increases.

The second and third conditions represent the self-confirming nature of those psychological factors. The parametric treatment of the psychological factors is similar in spirit to the formulation of the psychological games of Geanakoplos et. al. [12] and the fairness equilibrium of Rabin [18]. Before characterizing the morale equilibria, let us introduce a simplifying assumption to deal with the discreteness of effort level. As the utility function u_i is strictly concave in e_i, there would be a unique maximizer for u_i, if the efforts were continuous. Similarly, there would be a unique effort profile maximizing $\sum_i u_i$. When efforts are discrete, however, there may be two maximizers adjacent to the "true" maximizer in the continuous formulation. As this causes inessential complication in exposition, we exclude such a case. The necessary and sufficient conditions are the following.

Assumption: The cost function c is chosen generically so that $\Delta c(e)$ is not equal to 1, N, or $1 + K(\sum_{i=1}^{N}[m - e_i]_+)$ for any $e, e_1, ..., e_N, m \in \{0, 1, ..., L\}$.

Proposition 1 *All morale equilibria are symmetric and the set of morale equilibrium effort levels is* $E \equiv \{e|\ e^m \leq e \leq \overline{e}\}$, *where* e^m *is the Nash equilibrium with the material payoff, which is determined by*

$$\Delta c(e^m) < 1 < \Delta c(e^m + 1) \tag{2}$$

and $\overline{e}$ *is given by*

$$\Delta c(\overline{e}) < 1 + K(0) < \Delta c(\overline{e} + 1). \tag{3}$$

Proof. Any morale equilibrium is symmetric because for given k and m, each player i maximizes the same function

$$v(e_i) \equiv e_i - c(e_i) - k[m - e_i]_+. \tag{4}$$

Hence in any morale equilibrium $m = e^*$ and $k = K(0)$, where e^* is the symmetric effort level. As v is concave, e^* is an morale equilibrium effort level if the local incentive constraints $v(e^* - 1) < v(e^*)$ and $v(e^*) > v(e^* + 1)$ are satisfied. They are expressed respectively as

$$\Delta c(e^*) < 1 + K(0) \text{ and} \tag{5}$$

$$1 < \Delta c(e^* + 1). \tag{6}$$

Condition (2) identifies the smallest e^* to satisfy the latter, while (3) provides the largest e^* to satisfy the former. ■

Figure 1 shows a typical morale equilibrium e' and how the material Nash equilibrium e^m and the maximum morale equilibrium $\overline{e}$ are determined. For expositional simplicity, we ignore in the Figure the discreteness of effort and treat it as if it were a continuous variable. Note that the heavy line represents the marginal cost of *reducing* effort, while c' (corresponding to Δc in the discrete formulation) represents the marginal benefit.

Note that, with psychological payoffs, we have a "continuum" of equilibrium effort levels $e^m \leq e^* \leq \overline{e}$, each of which constitutes a *strict* equilibrium (an equilibrium where unilateral deviation strictly decreases one's payoff). Hence the traditional refinements concepts, such as perfectness, properness, or strategic stability, cannot tell which of them are most likely. We argue that the long run stochastic stability (Kandori, Mailath, and Rob [15] and Young [24]) is a natural concept to address stability of equilibrium in this model. Also note that the smallest morale equilibrium effort level corresponds to the "material" Nash equilibrium. Morale equilibrium effort level cannot be smaller, as providing more effort than the norm entails no psychological cost: at $e < e^m$, players want to increase their efforts. The maximum morale equilibrium effort level, $\overline{e}$, may be greater or smaller than the efficient effort level e^+, which maximizes the total material payoff $\sum_{i=1}^{N}(Ne_i - c(e_i))$. The "first order condition" is $\Delta c(e^+) < N < \Delta c(e^+ + 1)$. Comparing this with condition (3), we have:

Proposition 2 *The efficient effort level is attained iff* $N \leq 1 + K(0)$.

This is the first important property of norms and morale as resource allocation devices; their effectiveness depends on the relative strength of the psychological and material payoffs. The psychological factors can be *potentially* effective, if the material payoff is not overwhelming. In the next section, however, we show that high equilibrium norms and morale are dynamically unstable and identify what is sustainable in the long run. This is the second important property of norms and morale in resource allocation problems.

3 The Dynamics of Norms and Efforts

Let us now introduce the dynamics. At time t, player i's payoff is given by

$$u_i = \sum_{j=1}^{N} e_j(t) - c(e_i(t)) - k(t)[m(t) - e_i(t)]_+,$$

where

$$m(t) = \text{median}\{e_1(t-1), ..., e_N(t-1)\} \text{ and} \tag{7}$$

$$k(t) = K\left(\sum_{i=1}^{N}[m(t) - e_i(t-1)]_+\right). \tag{8}$$

(We assume that $m(0)$ and $k(0)$ are exogenously given.) In each period, each player maximizes the above payoff with probability $1-\epsilon$, and with probability ϵ he takes any effort level with equal probability. The latter eventuality is called *mutation*, and ϵ is referred to as the *mutation rate*. This defines a Markov chain with a finite state space where the state represents the current effort profile $\mathbf{e}(t)$.

Let us now consider the dynamic without mutation. At each time t, each player i maximizes

$$v_t(e_i) \equiv e_i - c(e_i) - k(t)[m(t) - e_i]_+.$$

As players maximizes the same function v_t, the strategy profile at t is symmetric for any given $\mathbf{e}(t-1)$. (Recall that we have assumed the unique maximizer of v_t, thanks to the Assumption.) Given any profile $\mathbf{e}(t-1)$, let us examine how the symmetric effort level e at time t is determined. (Case 1: $m(t) < e^m$) In this case, the player's payoff is maximized at $e > m(t)$. As only the material payoff matters for $e > m(t)$, we reach the material Nash equilibrium $e = e^m$. (Case 2: $m(t) \geq e^m$) Effort level e is equal to $m(t)$, when $\Delta c(m(t)) < 1 + k(t)$, as a downward deviation from the present norm $m(t)$ does not pay. Otherwise deviation $e < m(t)$ is beneficial and the optimal level of deviation is determined by $\Delta c(e) < 1 + k(t) < \Delta c(e+1)$. In any case, the new effort level e constitutes a morale equilibrium: As $0 \leq k(t) \leq K(0)$, e satisfies equilibrium conditions (5) and (6). We summarize what we have obtained as follows.

Proposition 3 *Suppose mutation rate ε is equal to 0. Given any $\mathbf{e}(t-1)$, $m(t)$, and $k(t)$, in the next period t players choose a morale equilibrium effort level $e(t)$, which is given as follows.*

(Case 1) $m(t) < e^m$: $e(t) = e^m$.

(Case 2) $m(t) > e^m$: If $\Delta c(m(t)) < 1 + k(t)$, then $e(t) = m(t)$. Otherwise $e(t) < m(t)$ and it is uniquely determined by

$$\Delta c(e(t)) < 1 + k(t) < \Delta c(e(t)+1).$$

Hence, the dynamics without mutation works as follows. If the median effort is lower than the material Nash effort e^m, then the material Nash equilibrium arises. Otherwise, the players reach a morale equilibrium whose effort level is smaller than or equal to the current median effort. This in particular implies that process without mutation always ends up with a morale equilibrium, which is formally stated as follows. (Recall that a *limit set* is an absorbing set of states under no mutation[9].)

Corollary 1 *Each morale equilibrium constitutes a limit set as a singleton, and there is no other limit set; the family of limit sets is* $\{\{\mathbf{e}\}|\mathbf{e} = (e, ..., e), e \in E\}$.

If random shocks (mutations) are present (i.e., if the mutation rate ϵ is positive), at each moment of time, each state is realized with a positive probability, and it is known that in such a case there is a unique stationary distribution, denoted $\mu(\epsilon)$. It represents the relative frequencies of states in the long run and is independent of the initial state. The support of $\mu^* = \lim_{\epsilon \to 0} \mu(\epsilon)$ is called the set of *long run stochastically stable (LSS) states*[10], and this is the set of states on which the system spends most of time in the long run, when the mutation rate is small but strictly positive. It is known that the set of LSS states corresponds to a collection of limit sets, and we call such a limit set (a limit set in the support of μ^*) a LSS limit set. We now identify the LSS limit sets by the *transition tree* technique developed by Freidlin and Wentzell [11], Kandori, Mailath, and Rob [15] and Young [24]. This approach considers trees defined on the family of limit sets and associated cost, and shows that the root of a minimum cost tree corresponds to a LSS limit set.

A transition tree is a directed graph, the set of whose nodes is equal to the family of all limit sets. Formally, it is a collection of directed branches between limit sets, where (i) there is one node, called the *root*, without an outgoing branch and (ii) any other node has a single outgoing branch, and (iii) there is no closed loop. Given Corollary 1, we abuse notation to say the "branch from e to e'" when we mean the "branch from limit set $\{\mathbf{e}\}$ to $\{\mathbf{e}'\}$", where $\mathbf{e} = (e, ..., e)$ and $\mathbf{e}' = (e', ..., e')$ are (symmetric) equilibria. We also say "morale equilibrium" or "state" e when we mean morale equilibrium or state $\mathbf{e} = (e, ..., e)$. With this convention in mind, we now identify the cost

[9] A set of states is a limit set if, under no mutation, (i) any two states in the set are mutually reachable (within a finite period) and (ii) no outside state is reachable from the states within the set.

[10] We follow this terminology proposed by Ellison [7].

of transition from equilibrium e to equilibrium e', denoted by $C(e, e')$. Let $(\mathbf{e}_0, \mathbf{e}_1, ..., \mathbf{e}_T)$ be a path in the state space from equilibrium $\mathbf{e}_0 = (e, ..., e)$ to another equilibrium $\mathbf{e}_T = (e', ..., e')$. Note that the intermediate states in the path $\mathbf{e}_t$ $(0 < t < T)$ do not have to be equilibria. Let $c(\mathbf{e}_{t-1}, \mathbf{e}_t)$ be the cost of transition (the number of required mutations) from state $\mathbf{e}_{t-1}$ to $\mathbf{e}_t$, and recall the definition of $C(e, e')$:

$$C(e, e') = \min \sum_{t=1}^{T} c(\mathbf{e}_{t-1}, \mathbf{e}_t), \tag{9}$$

where the minimum is taken over all paths $(\mathbf{e}_0, \mathbf{e}_1, ..., \mathbf{e}_T)$ such that $\mathbf{e}_0 = (e, ..., e)$ and $\mathbf{e}_T = (e', ..., e')$, for any $T = 1, 2, 3, ...$ The cost of a tree is the sum of the cost of branches, where the cost of branch from e to e' is defined by (9). The Freidlin-Wentzell transition tree analysis shows that a limit set is LSS if and only if it is the root of a minimum cost transition tree.

Now we determine the costs of "upward" transitions (from e to $e' > e$). In what follows the proofs of lemmas are given in the Appendix. First, let us show that upward transition requires that a *majority* (more than $N/2$) of players mutate. Recall that the number of players N is odd, so that $N/2$ is not an integer.

Lemma 1 *For any pair of morale equilibria $e < e'$, we have $C(e, e') > N/2$.*

The next lemma (and its proof) shows that equilibrium effort can "creep up" if more than the half of the players exert slightly more effort, provided that original effort level is less than a certain threshold (denoted e^U). Recall again that $\frac{N+1}{2}$ is the smallest integer which is more than $\frac{N}{2}$, as we assume that N is odd.

Lemma 2 *$C(e, e+1) = \frac{N+1}{2}$ if $e^m \leq e < e^U$, where e^U is the unique morale equilibrium effort level satisfying*

$$\Delta c(e^U) < 1 + K(\frac{N-1}{2}) < \Delta c(e^U + 1).$$

Note that e^U is close to the maximum equilibrium effort level $\overline{e}$, when the grid size for the effort level is sufficiently fine:

Remark 1 *If we introduce the grid size δ for effort level and suppose $e \in \{0, \delta, 2\delta, 3\delta, ...\}$, e^U is defined by $\Delta c(e^U) < 1 + K(\frac{N-1}{2}\delta) < \Delta c(e^U + 1)$. As the largest equilibrium effort $\overline{e}$ is defined by condition (3) $\Delta c(\overline{e}) < 1 + K(0) < \Delta c(\overline{e} + 1)$, we have*

$$e^U \to \overline{e}, \text{ as the grid size for effort } \delta \to 0.$$

Next we turn to identify the costs of "downward" transitions (from e to $e' < e$) . Let us first define effort level e^D to be the maximum effort level that satisfies

$$\Delta c(e^D) < 1 + K(\frac{N-1}{2}e^D). \tag{10}$$

This condition represents the incentives of the players when a minority of players shirk in the most effective way: i.e. when $\frac{N-1}{2}$ players deviate to 0. If this condition is satisfied, in the next period, everyone goes back to e^D, as the marginal benefit of reducing effort from e^D (the left hand side) is smaller than the marginal cost (the right hand side). Hence e^D *is the maximum equilibrium effort level that cannot be "pulled down" unless more than the half of the players reduce their efforts.* This turns out to be the maximum effort level sustainable in the long run; Theorem 1 below shows that the effort level between e^m and e^D are long run stochastically stable. By the definitions it is easy to see

$$e^m \leq e^D \leq e^U \leq \overline{e}.$$

Let us examine when the range of the long run sustainable effort levels is non-degenerate (in other words, when we have strict inequality $e^m < e^D$). As the material payoff is concave, a small deviation from the material Nash equilibrium entails minute cost, as Akerlof and Yellen [1] stressed. Hence there is a range of effort levels near the material Nash equilibrium, which are sustained by a modest binding power of the norm. The range is large when the material payoff is relatively flat near the equilibrium. Formally, recall that e^m is defined by $\Delta c(e^m) < 1 < \Delta c(e^m + 1)$, and note that $\Delta c(e^m) \cong 1 \cong \Delta c(e^m + 1)$ when the grid size for the effort level is sufficiently fine. In contrast, $K(\frac{N-1}{2}(e^m + 1))$ does not vanish as the grid size tends to zero, so that we have $\Delta c(e^m + 1) < 1 + K(\frac{N-1}{2}(e^m + 1))$. This means that $e^m < e^D$ when the grid size is fine, as e^D is the maximum effort level satisfying $\Delta c(e) < 1 + K(\frac{N-1}{2}e)$. Furthermore, if the material payoff is relatively flat at the material Nash equilibrium, $\Delta c(e) < 1 + K(\frac{N-1}{2}e)$ is satisfied for a wide range of e, and therefore the set of LSS effort level $\{e | e^m \leq e \leq e^D\}$ can be large. The next lemma briefly summarizes the above discussion.

Lemma 3 $e^m < e^D$ *if* $1 + K(\frac{N-1}{2}(e^m + 1)) > \Delta c(e^m + 1)$. *Otherwise,* $e^m = e^D$.

The next lemma formally shows that any morale equilibrium $e > e^D$ can be "pulled down" when less than the half of the players shirk.

Lemma 4 *For any morale equilibrium $e > e^D$, there is a morale equilibrium $e' < e$ such that $C(e, e') < N/2$.*

Although the above lemma is enough to prove our main result (Theorem 1 below), we can obtain a sharper characterization of the downwards costs, which helps to understand the nature of the dynamics.

Definition 2 *For $n = 1, 2, ..., \frac{N-1}{2}$, define effort level e^n to be the maximum effort level e satisfying*

$$\Delta c(e) < 1 + K(ne).$$

As Δc is increasing and K is decreasing, e^n is non-increasing in n, and note that $e^{\frac{N-1}{2}}$ is equal to e^D;

$$e^D = e^{\frac{N-1}{2}} \leq \cdots \leq e^2 \leq e^1.$$

Note that e^n is the maximum effort level that cannot be pulled down unless (strictly) more than n players shirk (exert zero effort). The reasoning is parallel to our discussion on e^D. Those effort levels are (approximately) given as in Figure 2, where we again assume for simplicity that e were a continuous variable so that c' in the figure plays the role of Δc. From Figure 2 we can see that the above inequalities are strict when the grid size on effort level (normalized as 1 in the current formulation) is sufficiently small. The next lemma provides the exact characterization of the downwards costs for $e > e^D$ (here e^0 is defined to be $\overline{e}$ for convenience).

Lemma 5 *For effort level e such that $e^n < e \leq e^{n-1}$, $n = 1, 2, ..., \frac{N-1}{2}$, we have*

$$\min_{e' \in E \setminus \{e\}} C(e, e') = n,$$

and the minimum cost is achieved by $e' \leq e^n$.

In other words, the most likely downward transition from e such that $e^n < e \leq e^{n-1}$ is to have downward mutation to zero effort level by n players, and this achieves a new equilibrium with a lower effort level $e' \leq e^n$. Note that, as the gain from downward deviation ($\Delta c(e)$) becomes larger as e increases (increasing marginal cost), the current equilibrium is upset by a small reduction of the binding power (i.e., by a relatively small number of (downward) mutations away from the current work norm e), when e is much larger than the material Nash effort level. As e becomes smaller and approaches the material Nash effort e^m, in contrast, the material gain

from downward deviation becomes smaller, as Akerlof and Yellen [1] noted. This means that the current equilibrium is upset only when a majority of the players shirk (in other words, only when the binding power becomes sufficiently small). This is the crux of the matter that determines the dynamics of norms and morale.

Hence we have shown that for any morale equilibrium $e > e^D$, a downward transition to a lower equilibrium effort level is possible when less than the half of the players shirk (i.e., with transition cost less than $N/2$). The next lemma shows that, for morale equilibrium $e \leq e^D$, the cost of a downward transition is always equal to $\frac{N+1}{2}$.

Lemma 6 *For any pair of morale equilibria e and e' such that $e' < e \leq e^D$, we have $C(e, e') = \frac{N+1}{2}$.*

Let us summarize in the following table the most likely transition from each state e, denoted by e' ($e' \in \text{Arg}\min_{e' \neq e} C(e, e')$), and its associated cost ($\min_{e' \neq e} C(e, e')$). This is obtained by the above Lemmas. In the table, it should be understood that $e^m \leq e'$. Note that e is displayed (from left to right) in the increasing order. (Hence, for example, the rightmost column indicates that the most likely transition from any state $e^1 < e \leq \overline{e}$ is to move towards $e' \leq e^1$, with associated cost 1.)

e	$e^m \cdots\cdots\cdots e^D$	$\cdots\ e^{\frac{N-1}{2}-1}$	$\cdots$	$\cdots e^2$	$\cdots e^1$	$\cdots \overline{e}$
e'	$e+1$ or[11] any $e' < e$	$e' \leq e^D$	$\cdots$	$e' \leq e^3$	$e' \leq e^2$	$e' \leq e^1$
cost	$\frac{N+1}{2}$	$\frac{N-1}{2}$	$\cdots$	3	2	1

Note that the states in

$$E^* \equiv \{e | e^m \leq e \leq e^D\} \tag{11}$$

are mutually reachable with cost $\frac{N+1}{2}$, and also note that this is the minimum cost of transition from each state in this set. We are now ready to state our main result; this "component" E^* corresponds to the set of the long run stochastically stable states:

Theorem 1 *The set of long run stochastically stable states is* $\{\mathbf{e} = (e, ..., e) \mid e^m \leq e \leq e^D\}$.

[11] There may be another $e' \geq e + 1$ that achieves the minimum cost.

Proof. By Corollary 1, we need to consider trees defined over the set of equilibria. Note that the following is true. For each node (equilibrium), endow an outgoing branch with the minimum cost, and then from the resulting graph delete a branch with the largest cost. If we obtain a tree, then it is a minimum cost tree. Let us first prove this assertion. Define, for each equilibrium e, the minimum cost of outgoing branch by

$$c^*(e) \equiv \min_{e' \in E \setminus \{e\}} C(e, e'),$$

and also define $c^{**} = \min_{e \in E} c^*(e)$. Take any tree and let us denote its root by e'. Its cost is at least $\sum_{e \in E} c^*(e) - c^*(e') \geq \sum_{e \in E} c^*(e) - c^{**}$, the cost of the tree constructed by the above procedure. Hence our assertion is proved. Let us now turn to the proof of the Theorem. Note first the following characterization of the minimum cost outgoing branches. (In what follows e and e' should be understood as equilibrium effort levels) For $e > e^D$, we can find $n < N/2$ such that $e^n < e \leq e^{n-1}$, and we have $c^*(e) = C(e, e^n) = n < N/2$ (by Lemmas 1 and 5). For $e \leq e^D$, $c^*(e) = \frac{N+1}{2}$ and $c^*(e) = C(e, e')$ if $e' < e$ or $e' = e + 1$ (by Lemmas 1,2, and 6). Then, we can construct a minimum cost tree whose root is any element e" in E^* as follows. In the first step choose the minimum cost branches in E^* as

$$e^m \to e^m + 1 \to \cdots \to e" - 1 \leftrightarrows e" \leftarrow \cdots \leftarrow e^D.$$

For $e \notin E^*$, (by Lemma 4) we can choose a minimum cost outgoing branch (e, e') such that $e' < e$ with cost less than $\frac{N}{2}$. Then delete the outgoing branch from e", which has the maximum cost $\frac{N+1}{2}$. The result is a tree with root e", because for each $e \notin E^*$, there is a path leading to E^*. From the above assertion this must be a minimum cost tree, and we conclude that any $e" \in E^*$ is a long run stochastically stable state. Note that the minimized cost of trees is equal to $\sum_{e \in E} c^*(e) - \frac{N+1}{2}$. Furthermore, there is no minimum cost tree whose root is $e" \in E \setminus E^*$. If so, the cost of the tree is at least $\sum_{e \in E} c^*(e) - c^*(e") > \sum_{e \in E} c^*(e) - \frac{N+1}{2}$, a contradiction. ■

The minimum cost tree constructed in the above proof and the above table suggest that the efforts gradually decay until the long run stochastically stable set E^* is reached, and in the (very) long run the efforts drift in this set. Simulation results confirm this observation. Figure 3 presents a sample path for a particular stage game, with 7 players and mutation rate $\epsilon = 0.15$. The LSS set E^* corresponds to the effort levels between 40 and 53, while the maximum equilibrium effort level is $\overline{e} = 80$. The median effort (the norm) gradually erodes until it hits the LSS set. Note that the erosion is

relatively quick. Figure 4 shows another sample path of the same model for a much longer time span. The median effort mainly drifts in the LSS set. A higher norm occasionally emerges, but it quickly erode until the LSS set is reached again.

4 Waiting Times

Let $E^n = \{e | e^m \leq e \leq e^n\}$ for $n = 1, 2, ..., \frac{N-1}{2}$. We now examine the expected time to reach this set of equilibrium effort levels. Note that E^n coincides with the set of the LSS states when $n = \frac{N-1}{2}$. This issue can be addressed by the notion of the *radius* and the *modified coradius*, powerful analytical tools developed by Ellison [7]. Here, we present a simpler, self-contained analysis, which is possible due to the special structure of the dynamics. In particular, the following fact greatly simplifies the analysis. From Proposition 3, we know that from any state $\mathbf{e}$, an equilibrium e^0 is reached *within one period*, if there is no mutation. Suppose that players 1 and 2 mutate to e_1 and e_2, when the current state is $\mathbf{e}$. The state in the next period is $(e_1, e_2, e^0, ..., e^0)$. The same state is obtained by the same eventuality if the current state were e^0. This observation shows the following lemma, where the one period transition probability from state $\mathbf{e}$ to $\mathbf{e}'$ is denoted by $p_\epsilon(\mathbf{e}, \mathbf{e}')$.

Lemma 7 *For any state* $\mathbf{e}$*, let* e^0 *be the equilibrium reached in the next period without mutation. Then, for any state* $\mathbf{e}'$*,* $p_\epsilon(\mathbf{e}, \mathbf{e}') = p_\epsilon(e^0, \mathbf{e}')$*.*

This lemma shows that it is sufficient to consider transitions from equilibria. The proof of Lemma 5 shows that, from any morale equilibrium $e \notin E^n$, an equilibrium $e' \leq e^n$ (hence E^n) is achieved if n players mutate to zero effort. Now let D^n be the set of states from which E^n is achieved with probability one, in the absence of mutation (the basin of attraction of E^n) and let $\sim D^n$ denote its complement. Lemma 7 implies that, to identify a lower bound of the one period transition probabilities form the states in $\sim D^n$ to the set D^n, we only need to consider the transition probabilities from the equilibria in $\sim D^n$ (i.e., $e \notin E^n$). The above argument shows that the probability that n players mutate to zero effort, denoted by p, serves as a lower bound[12]. Replace the one period transition probabilities from $\sim D^n$ to D^n with their lower bound p, and calculate the expected waiting time to reach from $\sim D^n$

[12]The probability of reaching E^n from $e \notin E^n$ in one period is larger than p, because having n mutations to zero is not the only way of reaching E^n.

to D^n, and denote it by W. (Note that, by construction, W is longer than the true waiting time.) Then, W satisfies $W = p \times 1 + (1-p)(1+W)$, or $W = 1/p$. As $p = A_n\epsilon^n + A_{n+1}\epsilon^{n+1} + \cdots + A_N\epsilon^N$, an upper bound of the expected waiting time to reach E^n is given by, for small enough ϵ,

$$W < A\epsilon^{-n},$$

for some constant $A > 0$[13]. This shows that, for small n, E^n is reached within a reasonable amount of time.

Let us now identify a lower bound of the expected waiting time to escape from E^n $(n > 1)$. For each state $\mathbf{e} \in D^n$, let $q(\mathbf{e})$ be the probability of escaping from D^n in one period. Lemma 1 shows that to reach an equilibrium $e' \notin E^n$ from $e \in E^n$, it takes at least $\frac{N+1}{2}$ mutations. This means that, when $\mathbf{e} \in E^n$, for sufficiently small ϵ, $q(\mathbf{e}) \leq B\epsilon^{\frac{N+1}{2}}$ for some constant $B > 0$[14]. Lemma 7 implies that this is also true for all $\mathbf{e} \in D^n$. Hence, if we suppose that the one period probability of escaping D^n were equal to its upper bound $B\epsilon^{\frac{N+1}{2}}$ for all $\mathbf{e} \in D^n$, the resulting expected waiting time, denoted W', is shorter than the true expected waiting time. As we have $W' = B\epsilon^{\frac{N+1}{2}} \times 1 + (1 - B\epsilon^{\frac{N+1}{2}})(1 + W')$, a lower bound of the expected waiting time to escape form D^n is given by

$$W' = A'\epsilon^{-\frac{N+1}{2}},$$

for $A' = 1/B > 0$. Let us now summarize what we have obtained. Recall that D^n is the basin of attraction of E^n, the interval of effort level between the material Nash effort e^m and e^n, and $\sim D^n$ is its complement. The "threshold" effort levels e^n for $n = 1, 2, ...\frac{N-1}{2}$ are determined as in Figure 2, where $e^m \leq e^D = e^{\frac{N-1}{2}} \leq \cdots \leq e^2 \leq e^1$. Also recall that $D^{\frac{N-1}{2}}$ coincides with the basin of attraction of the LSS states.

Theorem 2 *Let $W(\mathbf{e}, X, \epsilon)$ be the expected waiting time to reach a set of states X from state $\mathbf{e}$. Then, there are constants $A, A' > 0$ such that, for all sufficiently small ϵ, (i) $W(\mathbf{e}, D^n, \epsilon) \leq A\epsilon^{-n}$ for all $\mathbf{e} \in \sim D^n$, and (ii) $W(\mathbf{e}, \sim D^n, \epsilon) \geq A'\epsilon^{-\frac{N+1}{2}}$ for all $\mathbf{e} \in D^n$.*

[13] Choose A such that $AA_n > 1$. Then, $\frac{A\epsilon^{-n}}{W} \to AA_n > 1$, as $\epsilon \to 0$. In other words, $W < A\epsilon^{-n}$ for small enough ϵ.

[14] As escaping from D^n requires mutations more than or equal to $\frac{N+1}{2}$, we have $q(e) = \sum_{k=\frac{N+1}{2}}^{N} B_k(e)\epsilon^k$, where $B_k(e) \geq 0$. Choose B such that $B > B_{\frac{N+1}{2}}(e)$ for all $e \in E^n$. Then we have $\frac{q(e)}{B\epsilon^{\frac{N+1}{2}}} \to \frac{B_{\frac{N+1}{2}}(e)}{B} < 1$, as $\epsilon \to 0$. In other words, $q(e) < B\epsilon^{\frac{N+1}{2}}$ for sufficiently small ϵ.

A couple of remarks are in order. First, we have essentially the same waiting times if we replace D^n and $\sim D^n$ with E^n and $E \diagdown E^n$. This is because once $D^n \diagdown E^n$ is reached, E^n is reached in the next period (Proposition 3). Hence $W(\mathbf{e}, E^n, \epsilon) \leq W(\mathbf{e}, D^n, \epsilon) + 1$, and similarly, $W(\mathbf{e}, E \diagdown E^n, \epsilon) \leq W(\mathbf{e}, \sim D^n, \epsilon) + 1$. Secondly, and most importantly, the above Proposition indicates that efforts typically erode over time. It shows that the waiting time to reach E^n is in the order of ϵ^{-n}, while escaping E^n requires waiting time in the order of $\epsilon^{-\frac{N+1}{2}}$. As $n < \frac{N+1}{2}$, the latter is much longer than the former, for a small[15] ϵ. For example, suppose we have seven players and $\epsilon = 0.1$. If we start with the maximum morale equilibrium $\overline{e}$, the waiting time to reach an effort level lower than e^1 (or e^2) is in the order of $\epsilon^{-1} = 10$ (or $\epsilon^{-2} = 100$), but coming back requires a fairly large amount of time in the order of $\epsilon^{-4} = 10,000$. This means that the system gradually climbs down the ladder of equilibrium effort levels, and it is an example of what Ellison [7] called "step-by-step evolution". Such effects are observed in the simulation results in Figures 3 and 4. Third, as $E^{\frac{N-1}{2}}$ coincides with the LSS equilibria, the LSS may not be reached within a reasonable time span, when N is large (the waiting time is in the order of $\epsilon^{-\frac{N-1}{2}}$). However, our argument above shows that we do observe the effects of stochastic evolution; effort levels gradually erode, although they may not completely be reduced to the LSS effort levels in the relevant time span.

5 Concluding Remarks

In this section, we discuss related literature and provide a couple of remarks, some of which are highly speculative. The social dilemma game with norms and morale bears some similarity to the minimum effort game, in which player i's payoff is given by $\min\{e_1, ..., e_N\} - ce_i$, with $0 < c < 1$. Any symmetric effort profile is an equilibrium in this game. Van Huyck, Battalio, and Beil [22] reported experimental results showing that, with a large N, effort level converges to its minimum level (0). Their 1991 paper [23] considered the median game, where player i's payoff is given by $m - b(m - e_i)^2$, where m is the median of $\{e_1, ..., e_N\}$. They found that the subjects invariably converge to the equilibrium determined by the initial median. Note that, unlike our model, the median always has the same

[15] The two quantities n and $\frac{N+1}{2}$ correspond to the modified coradius and the radius of D^n, and Ellison's theorem [7] shows that the LSS states are *contained in* E^n, as the former is less than the latter, for each $n = 1, ..., \frac{N-1}{2}$. Our analysis in the previous section shows that all states in $E^{\frac{N-1}{2}} = E^*$ are in fact LSS states.

"binding power" b in their model, and there is no incentive to reduce e_i, even if the binding power b is equal to zero. Hence no erosion is likely to happen in their game. Crawford [5] presented a model with adaptive expectations of the minimum or median in those experiments to conduct econometric analysis. Anderson, Goeree, and Holt [2] and [3] presented a stochastic dynamic model for the minimum effort game and showed that a particular equilibrium is selected in the long run. Unlike our model, they postulate that (i) players rationally expects the current minimum effort distribution and (ii) the adjustment dynamic is in continuous time and subject to the shocks represented by a Brownian motion. Note that, in all of the works cited above, the material payoff itself has the coordination game structure, while in our model such a structure arises via the interaction of material and psychological payoffs. Isaac, Walker, and Williams [14] conjectured that, in their experiments where contributions to public good declined over time, the subjects maintained some cooperation as long as the current payoff is expected to be higher than the material Nash payoff, but no formal dynamic model was presented.

The dynamics of norms and standards have been explored in somewhat different contexts by Ellison [8], [9], and Sobel [21]. Ellison [8] provided detailed accounts for the slowdown of academic publishing[16]. In the companion paper [9] he considered a model where academic authors have two tasks, developing a new idea and its execution, the latter of which is relevant for the revision of a paper. The model shows that over time authors gradually put more efforts for revision, because of the persistent small misperception caused by the overconfidence in one's own work. Sobel [21] considered generations of players who wish to join a club. Each player exerts a multi-dimensional effort vector, whose components are aggregated into a one-dimensional performance index. One is accepted to the club if he is comparable to the current members, in terms of the index. Sobel showed quite generally that the fluctuations of the weights to compose the index result in declining standards. The multi-dimensional nature of effort is essential in those works, while our model is built on one-dimensional effort, where the reciprocal nature of altruism plays a major role.

Now we turn to some remarks.

1. Strategic Placement of Highly Motivated Workers: Relative strength of the material and psychological payoffs may vary across players.

[16] His empirical paper [8] indicates that a substantial part of the slowdown is caused by extensive revisions, and the delay of referee's reports accounts for a quarter of the slowdown. His model [9] sheds light on the former, while our work might be relevant for the latter.

Suppose that a firm consists of two factories, each of which is represented by our model with seven workers. Among the fourteen workers in the firm, six of them are highly motivated in the sense that they put higher weights on the psychological payoffs. In particular, assume that their weights of psychological payoffs are so high that they are willing to observe the norm, as long as a majority of fellow workers do so. Also suppose that this is the case even under random shocks (so that they do not mutate). How should one allocate the six good citizens? If we split them equally to each factory, a majority of workers in each factory behave as in our model. Then, the transition costs in Section 3 and the order of waiting times in Section 4 are unaffected, as they involve mutations by less than the half of the players. This means that *the basic properties of the dynamics of norms and morale are unaffected, unless we change the behavior of a majority of players.* Hence, the norm in each factory erodes in a similar way to our model, although the erosion is somewhat slower. In particular, effort level falls below e^3 in both factories in the long run. On the other hand, if we place all the six good citizens in one factory, we can sustain in that factory the maximum equilibrium effort level $\overline{e}$.

The past two decades witnessed the rigorous theoretical analysis of organizational design, based on game theory and economics of information. Material incentives are the key elements in such an approach, but we are left with the feeling that there are something other than incentives that matter in organization. The above tentative analysis suggests that taking psychological factors into consideration could be a fruitful way to go one step further. However, a caution is in order. Recently, a variety of anomalous behavior rules have been justified as stylized facts and introduced to economic analysis. We have to be cautious, however, to derive policy implications, unless we have enough information about the postulated behavioral rules. Recall that the early macro policy based on the stability of the Phillips curve, which was once conceived as a reliable stylized fact, spelled disaster. To derive dependable policy implications, we have to dig deeper into the black box of the anomalous behavior, as Rubinstein [20] argues.

2. A Catch in Economic Transition: Consider the situation where the present system sustains a certain payoff level, and suppose a new system, which is characterized by our social dilemma game, is proposed as an alternative, with the initial equilibrium effort being $\overline{e}$. If $\overline{e}$ induces a higher payoff than the status quo, people may adopt the new system, expecting that $\overline{e}$ is going to prevail. However, a gradual erosion may lead to a worse outcome than the status quo. To assess the merit of economic transition, one has to consider what is sustainable in the long run, where material

payoffs play a larger role.

3. Diversity near the Neoclassical Equilibrium: Our analysis shows that a certain range of effort levels can be sustained in the long run. As our simulation result (Figure 4) illustrates, the median effort (the work norm) changes in this range every once in a while. It is known that the relative frequencies of the norms in this time series data must coincide with those in the cross section data in the long run, as our system is ergodic. Hence, in our formulation, different organizations with the same material payoff structure typically exhibit different work norms, which persists for a long time. This happens in the vicinity of the Neoclassical equilibrium (i.e., the Nash equilibrium with respect to the material payoffs), where a small change of efforts entails a minute cost. Akerlof and Yellen [1] stressed that small deviations from optimization behavior can potentially explain a variety of stylized facts. Our model is in line with this research program, and it provides new insights into the diversity of performances of different organizations and economies.

Appendix

We present the proofs of the Lemmas here.

Proof of Lemma 1: We suppose $C(e, e') < N/2$ and show that this leads to a contradiction. Let $(\mathbf{e}_0, \mathbf{e}_{1,...,}\mathbf{e}_T)$ be the minimum cost path which achieves $C(e, e')$ and let me_t be the median effort level at $\mathbf{e}_t$. Let $BR(\mathbf{e}_t)$ be the (identical) effort level each player would exert under no mutation when previous state was $\mathbf{e}_t$. Proposition 3 shows how it is determined. We claim $me_{t+1} = BR(\mathbf{e}_t)$. Note that $C(e, e') < N/2$ implies that less than the half of the population mutate in the transition from $\mathbf{e}_t$ to $\mathbf{e}_{t+1}$ (i.e., $c(\mathbf{e}_t, \mathbf{e}_{t+1}) < N/2$). Hence, by Proposition 3, more than half of the population must be playing $BR(\mathbf{e}_t)$ at $\mathbf{e}_{t+1}$, so that the median effort level at $\mathbf{e}_{t+1}$ is equal to $BR(\mathbf{e}_t)$, as claimed. We note that for all $t = 0, ..., T$, me_t is a morale equilibrium effort level. The reason is twofold; (i) $me_t = BR(\mathbf{e}_{t-1})$, for $t = 1, ..., T$ and $BR(\mathbf{e}_{t-1})$ is always a morale equilibrium effort level (by Proposition 3), and (ii) $m_0 = e$ is a morale equilibrium effort level by definition. Hence $me_t \geq e^m$, as e^m is the smallest morale equilibrium effort. Therefore, Case 2 of Proposition 3 applies and we have $me_t \geq BR(\mathbf{e}_t)$. This and the former claim shows $e = me_0 \geq me_1 \geq ... \geq me_T = e'$, which contradicts $e < e'$.

Proof of Lemma 2: Suppose that, at equilibrium e, $\frac{N+1}{2}$ players' effort levels mutate towards $e + 1$. The median effort in the new state is $e + 1$,

and the $\frac{N-1}{2}$ non-mutants are deviating downwards (by 1) from this new norm. Hence the binding power of the new norm is $K(\frac{N-1}{2} \times 1)$. By Case 2 in Proposition 3 and $\Delta c(e) < \Delta c(e^U) < 1 + K(\frac{N-1}{2})$ (recall that Δc is increasing), in the following period equilibrium $e+1$ is achieved. By Lemma 1, this is the minimum cost path.

Proof of Lemma 4: We show that there is a path from e to e' with cost less than $N/2$. Let c be the minimum integer satisfying

$$\Delta c(e) > 1 + K(ce)$$

By the definition of e^D, we must have[17] $\Delta c(e) > 1 + K(\frac{N-1}{2}e)$ for any $e^D < e$. As K is decreasing, we have $c \leq \frac{N-1}{2}$ $(< N/2)$. Now suppose that, at equilibrium e, c players mutate into effort level 0. As $c < N/2$, the median at the new state remains to be e, so that the sum of downward deviations is equal to ce. Then, the displayed inequality above and Case 2 of Proposition 3 show that in the following state all players take a morale equilibrium action $e' < e$. As the cost of this path is equal to c, we have $C(e, e') \leq c < N/2$.

Proof of Lemma 5: Suppose n players mutate to zero effort. In the next period, all players choose effort level e' defined by

$$\Delta c(e') < 1 + K(ne) < \Delta c(e' + 1).$$

(See Case 2 of Proposition 3.) As e^n is defined to be the maximum effort level satisfying $\Delta c(e) < 1 + K(ne)$, we must have

$$\Delta c(e^n + 1) \geq 1 + K(n(e^n + 1)).$$

As $e \geq e^n + 1$ and $K(\cdot)$ is decreasing, we have

$$\Delta c(e^n + 1) \geq 1 + K(n(e^n + 1)) \geq 1 + K(ne) > \Delta c(e').$$

Since Δc is increasing, we conclude $e^n + 1 > e'$, or equivalently, $e^n \geq e'$. Hence $e' \leq e^n$ is achieved with cost n. As $n < N/2$, there is no upward transition with the same cost (Lemma 1). Suppose there is $e' < e$ such that $C(e, e') < n$. This leads us to a contradiction. Let $(\mathbf{e}_0, ..., \mathbf{e}_T)$ be the minimum cost path which achieves $C(e, e')$ and let me_t be the median effort level at $\mathbf{e}_t$. Let $BR(\mathbf{e}_t)$ be the best reply effort level to $\mathbf{e}_t$. As

[17] Note that, thanks to the Assumption, there is no integer e that satisfies $\Delta c(e) = 1 + K(\frac{N-1}{2}e)$.

$C(e,e') < n < N/2$ implies that a majority of players are always taking the best reply to the previous state, we have $me_{t+1} = BR(\mathbf{e}_t)$, as in the proof of Lemma 1. Now we argue $me_t = e$ for all $t = 0, ..., T$. (This contradicts the requirement $me_T = e' < e$.) The claim is shown by induction. As the claim is true for $t = 0$ by definition ($me_0 = e$), let us suppose $me_t = e$ is true and show $me_{t+1} = e$, for any $t = 0, ..., T-1$. The norm at $t+1$, defined to be me_t, is equal to e by the induction hypothesis. We now show that the binding power of this norm is sufficiently strong to deter deviations. Recall that $c(\mathbf{e}_t, \mathbf{e}_{t+1})$ is the number of mutations in the transition from $\mathbf{e}_t$ to $\mathbf{e}_{t+1}$ and $c(\mathbf{e}_t, \mathbf{e}_{t+1}) \leq n-1$ (as $c(\mathbf{e}_t, \mathbf{e}_{t+1}) \leq C(e,e') < n$). Note that the sum of downward deviations from the norm at t is maximized (and therefore the binding power is the weakest) when the mutants take effort level 0, and the maximum total downward deviation is equal to $c(\mathbf{e}_t, \mathbf{e}_{t+1})$ (the number of mutants) times $(e-0)$ (the maximum deviation). Hence, if we denote the binding power of the current norm e by k_{t+1}, we have $k_{t+1} \geq K(c(\mathbf{e}_t, \mathbf{e}_{t+1})e)$, as K is decreasing. By the definition of e^{n-1} (see Definition 2) and $e \leq e^{n-1}$, together with $c(\mathbf{e}_t, \mathbf{e}_{t+1}) \leq n-1$ and the fact that Δc is increasing and K is decreasing, we have $\Delta c(e) \leq \Delta c(e^{n-1}) < 1 + K((n-1)e^{n-1}) \leq 1 + K(c(\mathbf{e}_t, \mathbf{e}_{t+1})e) \leq 1 + k_{t+1}$, or

$$\Delta c(e) < 1 + k_{t+1}.$$

Case 2 in Proposition 3 then shows that $BR(\mathbf{e}_t) = e$. As we have shown $me_{t+1} = BR(\mathbf{e}_t)$, the proof by induction is completed.

Proof of Lemma 6: The proof consists of two parts. First we prove $C(e,e') \geq \frac{N+1}{2}$, and then we show that generally for any $e' < e$ (e does not have to be less than or equal to e^D) there is a path from e to e' with cost $\frac{N+1}{2}$.

Suppose, on the contrary, $C(e,e') < N/2$. This leads to a contradiction. The proof is similar to that of Lemma 5 and therefore omitted. Now suppose, at equilibrium e, $\frac{N+1}{2}$ players mutate into a lower equilibrium effort level $e' < e$. Then, at the new state, the median effort is e', and there is no downward deviation from e'. Hence the best reply at this state is equal to the best reply at equilibrium e' (as each player maximizes (4) with $m(t) = e'$ and $k(t) = K(0)$ both at this state and at equilibrium e'.). Therefore, in the following state all players take effort level e'. The cost of this path from e to e' is equal to $\frac{N+1}{2}$.

References

[1] Akerlof, G. and J. Yellen (1985) "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?", *American Economic Review*, Vol. 76, pp. 708-720.

[2] Anderson, S., J. Goeree, and C. Holt (1999) "Stochastic Game Theory: Adjustment to Equilibrium Under Noisy Directional Learning", mimeo., University of Virginia.

[3] Anderson, S., J. Goeree, and C. Holt (2001) "Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium", *Games and Economic Behavior*, Vol. 34, pp. 177-199.

[4] Andreoni, J. (1988) "Why Free Ride? Strategies and Learning in Public Goods Experiments", *Journal of Public Economics*, Vol. 37, pp. 291-304.

[5] Crawford, V. (1995) "Adaptive Dynamics in Coordination Games", *Econometrica*, Vol. 63, pp.103-143.

[6] Dawes, R. and R. Thaler (1988) "Cooperation", *Journal of Economic Perspectives*, Vol. 2, pp. 187-197.

[7] Ellison, G. (2000) "Basin of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution", *Review of Economic Studies*, Vol. 67, pp. 17-45.

[8] Ellison, G. (2001) "The Slowdown of the Economics Publishing Process", forthcoming in *Journal of Political Economy.*

[9] Ellison, G. (2001) "Evolving Standards for Academic Publishing: A *q-r* Theory", forthcoming in *Journal of Political Economy.*

[10] Fehr, E. and K. Schmidt (1999) "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics*, Vol. CXIV, pp. 817-868.

[11] Freidlin, M. and A. Wentzell (1984) *Random Perturbations of Dynamical Systems*, New York: Springer-Verlag.

[12] Geanakoplos, J., D. Pearce, and E. Stacchetti (1989) "Psychological Games and Sequential Rationality", *Games and Economic Behavior*, Vol. 1, pp. 60-79.

[13] Isaac, M., K. McCue, and C. Plott (1985) "Public Goods Provision in an Experimental Environment", *Journal of Public Economics*, Vol. 26, pp. 51-74.

[14] Isaac, M., J. Walker, and A. Williams (1994) "The Group Size and the Voluntary Provision of Public Goods", *Journal of Public Economics*, Vol. 54, pp. 1-36.

[15] Kandori, M., G. Mailath, and R. Rob (1993) "Learning, Mutation, and Long Run Equilibria in Games", *Econometrica*, Vol. 61, pp.29-56.

[16] Levine, D. (1998) "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics*, Vol. 1, pp. 593-622.

[17] Lindbeck, A., S. Nyberg, and J. Weibull (1999) "Social Norms and Economic Incentives in the Welfare State", *Quarterly Journal of Economics*, Vol. 144, pp. 1-35.

[18] Rabin, M. (1993) "Incorporating Fairness into Game Theory", *American Economic Review*, Vol. 83, pp. 1281-1320.

[19] Rabin, M. (1998) "Psychology and Economics", *Journal of Economic Literature*, Vol. XXXVI, pp. 11-46.

[20] Rubinstein, A. (2001) "Is it "Economics and Psychology"?: The Case of Hyperbolic Discounting", *NAJ Economics*, Vol. 1.

[21] Sobel, J. (2000) "A Model of Declining Standards", *International Economic Review*, Vol. 41, pp. 295-303.

[22] Van Huyck, J., R. Battalio, and R. Beil (1990) "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure", *American Economic Review*, Vol.80, pp. 234-248.

[23] Van Huyck, J., R. Battalio, and R. Beil (1991) "Strategic Uncertainty, Equilibrium Selection Principles, and Coordination Failure in Average Opinion Games", *Quarterly Journal of Economics*, Vol. 106, pp. 885-910.

[24] Young, P. (1993) "The Evolution of Conventions", *Econometrica*, Vol. 61, pp. 57-84.

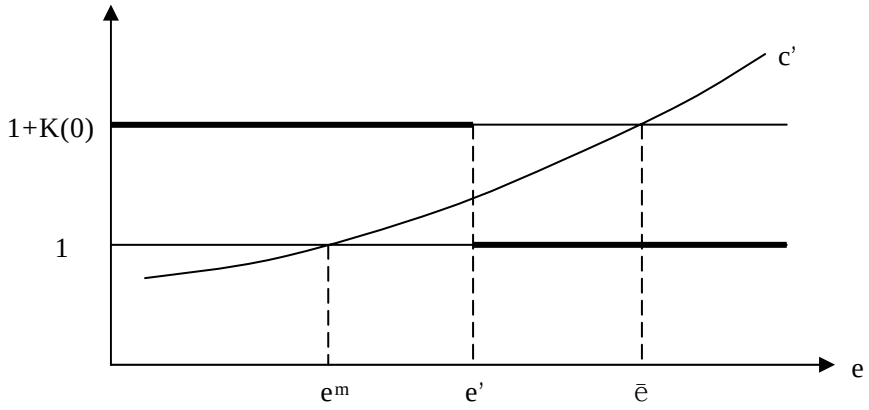

Figure 1.

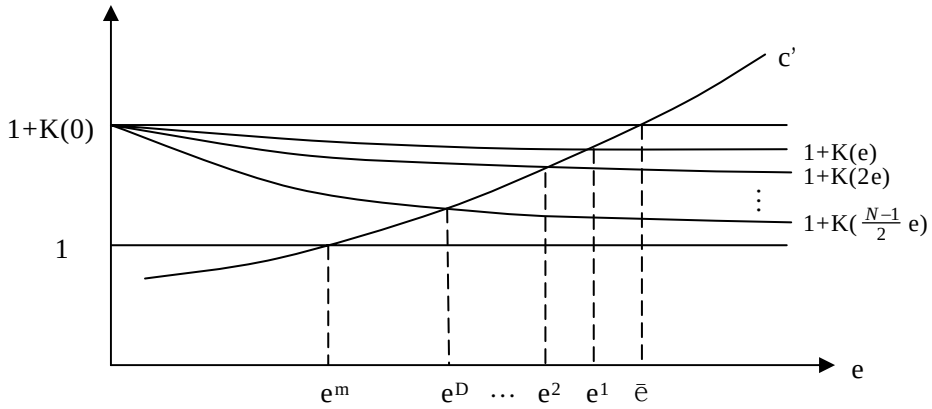

Figure 2.

Figure 3: c(e)=(e^2)/80, K(D)=80/(80+D), N=7, mutation rate=0.15

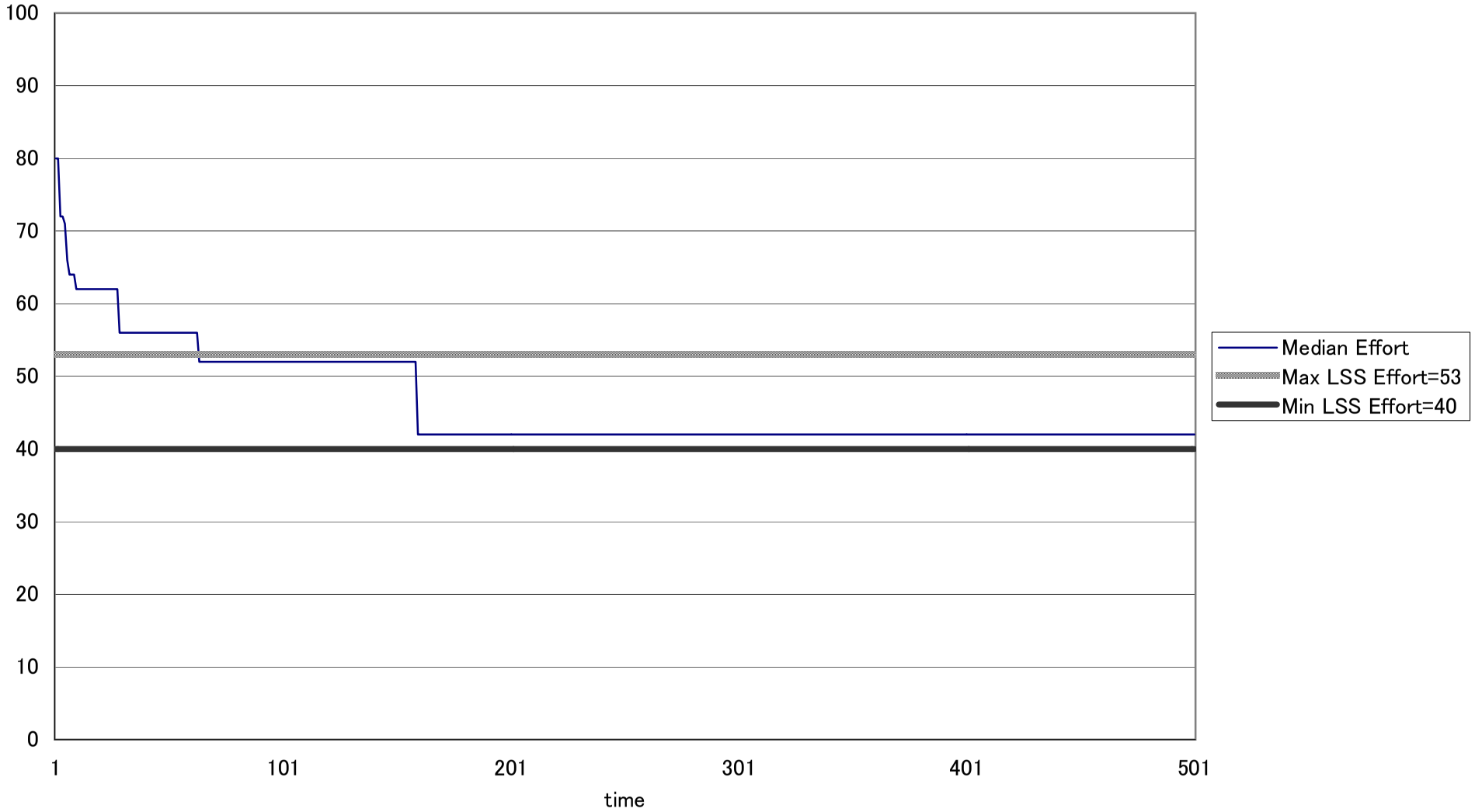

Figure 4: c(e)=(e^2)/80, K(D)=80/(80+D), N=7, mutation rate=0.15