Sustaining Cooperation through Strategic Self-Interested Actions*

Francesco Lancia  Alessia Russo

February, 2017

Abstract

This paper studies how organizations might promote cooperation between their members when individual contributions to an organization’s output are imperfectly observable. It considers an overlapping-generations game in which members with conflicting interests expend effort in pursuing self-interested outside tasks that are perfectly observable, in addition to the effort devoted to increasing the organization’s output. We show that it is optimal for both the organization and its members to strategically link the extent of cooperation to the effort expended on outside tasks. In the resulting equilibrium, the self-interested effort is distorted in order to signal members’ willingness to cooperate. After extending the game to multiple generations, we apply it to optimal task assignment along career paths in an organization.

JEL Classification: C73, D62, M54.

Keywords: Career Path, Common Good, Intergenerational Cooperation, Organizations, Public Perfect Equilibrium.

*Francesco Lancia, University of Vienna, email: francesco.lancia@univie.ac.at. Alessia Russo, BI Norwegian Business School, email: alessia.russo@bi.no. We thank Daniel Garcia, Bård Harstad, David K. Levine, Anirban Mitra, Nicola Pavoni, Nicola Persico, Debraj Ray, Karl Schlag, Paolo Siconolfi, and Timothy Worrall for helpful suggestions. We also benefitted from comments made by participants in the 2nd Personnel Economics and Public Finance conference in Ravello, the 11th SAET meeting in Faro, the 17th ISNIE conference in Firenze, and the seminars at the EIEF and the universities of Amsterdam, Bergen, Bologna, Firenze, Innsbruck, Modena, Napoli, Oslo, and Vienna. We also acknowledge the support of an EIEF research grant and the European Research Council under the EU’s 7th Framework Programme, ERC GA no. 283236.
1 Introduction

Public agencies and private firms are classic examples of organizations which over time recruit junior members to replace senior ones who are retiring from the organization. When they join the organization, new members become recipients of, as well as contributors to, the organization’s output. The success of these organizations therefore depends on their members’ willingness to refrain from individually profitable actions and to work for the common good. The design of pecuniary incentive schemes constitutes the standard economic approach to this problem.\(^1\) However, at a time when firms are cutting back on their financial incentive programs, as they are in most countries, non-pecuniary motivators may be an alternative way of building long-term engagement among members.\(^2\)

The model presented here shows how the motivation to contribute to the organization’s output can be aligned with the desire to build up an individual’s reputation.

Ever since the seminal contribution of Hammond (1975), a vast literature has developed to analyze the sustainability of cooperation based on reputation in ongoing organizations. In this literature, cooperation can be sustained as an equilibrium by means of a chain of rewards and penalties spanning over generations (see, e.g., Cremer, 1986; Kandori, 1992; Kreps, 1996). This literature builds on the idea that collective decisions are perfectly observable and that members expend effort along one dimension.\(^3\) In the real world, however, contributions are largely unobservable. Cooperative relationships are thus typically plagued with free-rider problems and private and social optima do not coincide. Moreover, members can expend effort along several dimensions, some of which are only privately beneficial. For instance, tenured managers and employees within a firm, collaborators in a research team or physicians in countries with a dual health system can devote part of their physical and mental resources to activities outside the organization, such as consulting activities for other organizations or training activities that provide them with additional qualifications or skills.\(^4\)

\(^1\)See Besley and Ghatak (2015) for an overview of standard models of agency problems.

\(^2\)In a study by Hay Group of 2000 organizations from 88 countries for the year 2009, 36% of the organizations surveyed reported that they were currently freezing pay, and 21% stated that they had either already reduced incentives or that they were considering reducing them in the future. The survey is available at www.haygroup.com/en/our-services/engage/employee-effectiveness-surveys. Non-monetary incentives, however, have been shown to improve intrinsic motivations more than monetary incentives. In a survey by McKinsey & Company of 1047 executives, managers, and employees for the year 2010, the respondents viewed non-cash motivators, such as praise and commendations, leadership attention and chances to lead projects, as being more effective than common monetary incentives, such as cash bonuses, increased base pay, and stock options. The survey is available at www.mckinsey.com/business-functions/organization/our-insights/motivating-people-getting-beyond-money.

\(^3\)See Salant (1991) and Smith (1992) for folk theorems which assume near-perfect patience when generations overlap for a sufficiently long time.

\(^4\)In some professions outside activities are explicitly allowed, while in others workers’ tasks are regulated to address potential conflicts of interest. This tension, as well as the potential gains, arising from
The effect of interactions among multiple activities determines the power of incentives to achieve the common good. Such interactions depend on whether the activities are substitutes or complements. The research questions are therefore: Is it desirable for the organization to adopt a strategy that links the extent of cooperation to a privately beneficial effort when the details of cooperation between members are imperfectly observable? And is this strategy also to the benefit of the organization’s members?

To address these questions, we consider a repeated game à la Cremer (1986). The basic structure is that of an organization whose members belong to two different generations at any point in time: the junior member who joins the organization in the current period and the senior member who joined it in the previous period and who will exit before the next one. While the two generations have different time horizons, they share the same preferences and face the same strategic possibilities in each period. Within each period, members interact via a prisoner’s dilemma type of game in which they simultaneously choose how much effort to devote to producing the organization’s output. Total output depends on the sum of their efforts. We depart from Cremer (1986)’s setup in two main respects: First, we assume that the total output only provides noisy information about individual contributions. Second, we allow members to also devote effort to tasks outside the organization, which are assumed to be perfectly observable and to benefit the agent alone. When such a self-interested effort exhibits complementarity or substitutability with the collective decision, we refer to it as a self-commitment action.

We focus on the Public Perfect Equilibrium (hereafter PPE) and study the best rule for promoting cooperative behavior among the organization’s members. We distinguish two regimes which differ in their history-contingent strategies: In the first, which we refer to as individual-based, the history of self-commitment actions does not align individuals’ behavior since members ignore this type of signal. In the second, which we refer to as organization-based, the history of self-commitment actions is instead part of the relevant history. It follows that while in the former regime self-commitment actions are taken as individual static best responses, in the latter, dynamic incentives arise and self-commitment actions involve a strategic choice to manipulate the cooperative response of future generations.

Two main predictions emerge from comparing the best PPE for each of the two regimes: First, the organization’s performance is higher in the organization-based regime than in the individual-based one. The intuition behind this result is as follows: When the effort for the common good is imperfectly observable, agents are always tempted to shirk.
To deter deviation, the best strategy entails forgiving potential opportunistic behavior with some probability. Suppose now that members can devote effort to tasks outside the organization and that the generated signal is not ignored by the other members. When such an effort is observable, the best strategy involves triggering reversion to autarky with certainty in the event of deviation from the agreed-upon level of outside tasks. In the presence of interaction between privately and collectively beneficial efforts, the best strategy also requires members to distort their performance in outside tasks relative to their performance if cooperation were fully enforceable. By doing so, agents endogenously undermine their short-term gains from opportunistic behavior. All members in turn correctly internalize the changes in individual incentives and accordingly soften their retaliatory responses. Ultimately, the strategy increases forgiveness so that cooperation is more likely in the long run and the organization’s expected output is larger.

The second prediction states that it is always worthwhile for members to comply with strategies that link privately and collectively beneficial efforts. Junior members face an intertemporal tradeoff associated with being in the organization-based regime rather than in the individual-based one. On the one hand, they are compelled to sustain a short-term loss due to the distortion of the privately beneficial effort needed to signal their willingness to cooperate. On the other hand, they benefit from the long-term gain resulting from the higher level of forgiveness. We show that the individual long-term gain is always worth the short-term loss regardless of the degree of monitoring imperfection. This is because in the organization-based regime members exert the optimal level of privately beneficial effort while internalizing its dynamic impact on future members’ equilibrium play.

We also show that formal institutions, such as, for example, mandatory provision of activities for the common good or rules establishing a minimum amount of working time or tasks in the organization, undermine rather than promote the self-enforcement of cooperation. At first glance, this appears to contradict conventional wisdom. Although the imposition of a mandatory provision constraint certainly reduces the individual gains from deviation and, in turn, increases the organization’s expected output, it has a negative effect on the individual incentives to comply with cooperation since it weakens the strategic role of self-commitment actions. Indeed, a tighter mandatory provision constraint has a twofold effect: First, it increases the marginal cost associated with exerting a distorted level of self-interested actions, and second, it lowers the marginal benefits from more forgiveness.

Although the baseline model is quite stylized, the strategic role of self-commitment actions survives even with the addition of more realistic features. For instance, the optimal level of self-commitment actions may vary with the size of the organization or the seniority of its members. We show how results generalize to an environment with
more than two generations and provide implications in terms of optimal task assignment along members’ career paths within an organization. Specifically, requiring the youngest members to build up their reputations by means of outside tasks is the best practice for the organization to adopt when members are overly impatient. This is because the junior member’s expected utility is low and in turn individual incentives to work for the common good are weak. Over-performing in outside activities that are complements to inside activities, or under-performing when they are substitutes can boost motivation to cooperate among members.\(^5\) This extension can be used to deepen our understanding of the reasons why generations may be differentially committed to the organization and how to improve the workplace engagement of all members.\(^6\)

The paper draws from the literature on strategic interactions with multiple actions initiated by Spence (1977) and Dixit (1980) and developed by Fudenberg and Tirole (1984) and Benoit and Krishna (1987), among others. The idea they put forward is that harsher punishments should improve incentives to cooperate by reducing the value of the players’ outside option. This is the logic of deterrence, according to which an agent can invest preemptively in technology if such investments make it more costly for other agents to deviate. In contrast to this literature, we consider multiple actions that are interdependent.\(^7\) This leads to a new mechanism involving self-commitment actions, which facilitates cooperation by endogenously undermining the short-run gain from deviation rather than by varying the retaliatory power of the punishment scheme.\(^8\)

The paper is therefore also related to the literature on issues linkage which uses supermodular games with multiple actions to study collusion in oligopolies with multi-market contracts (see, e.g., Bernheim and Whiston, 1998; Spagnolo, 1999; Cooper and Ross, 2009). In these models, agents interact simultaneously on several issues, all characterized by the strategic structure of a prisoner’s dilemma type of game. When issues are substitutes, linking them makes punishment harsher and deviation less worthwhile. The

---

\(^5\)This result is related to Shepsle and Nalebuff (1990) who adapt Cremer (1986) to show how a seniority system may help a political party to self-enforce cooperation by front-loading cooperative effort, namely, requiring junior members to pay their dues while senior-most members fully enjoy the perks of office. Unlike the model presented here, they do not allow members to exert self-interested actions nor characterize the best equilibrium.

\(^6\)According to the study “How Millennials Want to Work and Live” by Gallup Inc. (2016), workforce engagement differs between young and old generation employees, in that only 29% of millennials are engaged at work. Moreover, younger workers prefer more flexible working time compared to older ones who instead prefer a more traditional job model including visibility in the office during certain hours. The report is available at www.gallup.com/businessjournal/195209/few-millennials-engaged-work.aspx.

\(^7\)See Topkis (1998) for a review of the theory of supermodularity and complementarity.

\(^8\)The idea that technology investments can affect ex-post effort constraints is also present in the relational contracting literature. Ramey and Watson (1997) explore this idea in a model with repeated trading, where the principal can make a noncontractible irreversible investment before trade starts. We depart from that model since we do not consider infinitely-lived agents nor perfect information about cooperative efforts. Moreover, we allow agents to invest in every period.
opposite holds when issues are complements. We depart from these models by allowing agents to take a self-interested action when playing a prisoner’s dilemma game. We find that distorting privately beneficial effort and strategically linking it with the common good can increase cooperation whether decisions are substitutes or complements.\(^9\)

Methodologically, the paper is related to the literature on repeated games with imperfect public monitoring, which primarily deals with optimal penal codes (see, e.g., Green and Porter, 1984; Radner, 1986; Abreu, Pearce, and Stacchetti, 1990; Abreu, Milgrom, and Pearce, 1991). Agents’ inability to detect other players’ opportunistic behavior with certainty results in inefficient punishments. Fudenberg, Levine, and Maskin (1994) identify the conditions on the information structure which ensure that all feasible and individually rational payoffs can be supported in equilibrium. These results, however, do not hold when players are short-lived.\(^10\) In this context, we show that cooperation increases when agents exert privately beneficial efforts in order to signal their willingness to cooperate.

Finally, the paper is connected to the multitasking and job design literature (see, e.g., Holmstrom and Milgrom, 1991 and 1994; Baker, Gibbons, and Murphy, 1994). When different tasks have heterogeneously measurable outcomes, this literature shows that offering stronger incentives to perform tasks that are more measurable distorts effort away from other tasks. The main implication in terms of optimal allocation of tasks is that less discretion should be given to a manager in pursuing outside activities when inside effort is harder to measure. In contrast, we show that building up one’s reputation through both inside and outside activities motivates members to increase their performance within the organization.

The paper proceeds as follows: Section 2 presents the model’s setup and discusses benchmark results. Section 3 derives the best PPE in each of the two regimes. Section 4 compares the two equilibrium outcomes and discusses the effect of introducing mandatory provision constraints. Section 5 extends the results to include more than two generations. Section 6 concludes. The proofs that are not in the text can be found in the Appendix.

\(^9\)Acemoglu and Jackson (2015) use an overlapping-generations game to study how a mechanism that specifies strategies contingent on the history of an action that is visible to all future agents can facilitate cooperation. Their model, however, differs from ours in several respects: First, observable actions are adopted by a leader who can influence the cooperative behavior of his followers. Second, the interactions between generations occur via a coordination game. Finally, there is no interdependence between observable and unobservable actions.

\(^10\)Bhaskar (1998) shows that in an overlapping-generation context no cooperation is the unique equilibrium in pure strategies when only a finite number of periods of past play can be observed. With no memory of previous play, Lagunoff and Matsui (2004) and Anderlini, Gerardi, and Lagunoff (2008) show that if agents are altruistic and can send messages, then a folk theorem holds.
2 The Setup

Time is discrete and indexed by \( t = 0, 1, \ldots \). The model consists of an ongoing organization with an overlapping generations demographic structure, whose members live for two periods. Each generation is composed of one single member.\(^{11}\) At each time \( t \), a new member \( i \) enters the organization. She is young, denoted by \( y \), in the first period and old, denoted by \( o \), in the second period.

The organization’s members make choices in both periods of life and can exert effort along two dimensions: a privately beneficial effort and a collectively beneficial effort, which differ in both economic scope and observability. First, members decide on their privately beneficial effort, denoted by \( b_i^t \in \mathbb{R}_+ \), which is perfectly observable by both currently living and future generations. After this decision has been made for the current period, members choose a level of collectively beneficial effort, denoted by \( a_i^t \in \{a, \bar{a}\} \) with \( 0 \leq a < \bar{a} \), which is not observable by other members.\(^{12}\) Choosing \( \bar{a} \) is interpreted as cooperation, whereas choosing \( a \) is interpreted as shirking. This effort goes toward producing the organization’s output, which can be thought of as a common good consumed by all members. It is denoted by \( g_t = G(\sum_i a_i^t) + \epsilon \) where \( G(\cdot) \) is an increasing function of individual effort and \( \epsilon \) is an i.i.d. random variable with zero mean and cdf \( F(\epsilon) \). Unlike the action \( a_i^t \), the realized output \( g_t \) is perfectly observable by all generations. Exerting effort is costly, where the total cost \( C(a_i^t, b_i^t) \) borne by each member is strictly increasing in both types of effort and strictly convex in \( b_i^t \).

Each member derives utility from the organization’s output as well as from her privately beneficial action. We assume quasi-linear preferences. Hence, the per-period utility of member \( i \) at time \( t \) can be written as:

\[
\lambda g_t + \theta b_i^t - C(a_i^t, b_i^t),
\]

where \( \lambda > 0 \) captures the preference weight on the organization’s output and \( \theta > 0 \) measures the marginal benefit of the privately beneficial action.

The organization’s output is clearly maximized when both generations cooperate. However, we assume that at each time \( t \) member \( i \) prefers to shirk rather than cooperate for any level of collectively beneficial effort expended by the other member, denoted by

---

\(^{11}\)The assumption that there is one agent in each generation simplifies the analysis, but is not essential to the argument. Allowing for multiple agents within each generation would change the intratemporal incentive structure without modifying the intertemporal tradeoff that is the focus of the analysis.

\(^{12}\)The analysis remains unaltered if privately and collectively beneficial efforts are taken simultaneously when the organization is inhabited by two generations in each period.
as well as for any level of privately beneficial effort chosen by her:

\[ \lambda \left( G(a + a_i^{-i}) - G(a + a_i^{-i}) \right) < C(a, b_i^i) - C(a, b_i^i). \]

No cooperation is therefore the organization’s outcome in the absence of institutions that provide the necessary incentives to exert a high level of effort.

When the per-period utility of each member exhibits interdependence between collectively and privately beneficial efforts, a higher or lower level of effort along one dimension can have an impact on the individual’s incentives to exert effort in the other dimension. Let \( \Delta(b_i^i) := C_b(a, b_i^i) - C_b(a, b_i^i) \) measure the marginal impact of variation in \( b_i^i \) on member \( i \)'s gain from deviating from cooperative behavior. Self-commitment actions are then defined as self-interested actions that satisfy the following property:

**Definition 1.** A self-commitment action is any \( b \) such that \( \Delta(b) \neq 0 \).

The efforts can either be complements, i.e., \( \Delta(b) < 0 \), or substitutes, i.e., \( \Delta(b) > 0 \). In the former case, the incremental gain from choosing \( a \), i.e., the benefit from shirking, decreases as \( b \) increases. The opposite holds in the case of substitutability.\(^\text{13}\) These types of interdependence between efforts are commonly observed in organizations. As an example of substitutability, employees can devote part of their physical and mental resources to private activities outside the organization, such as, for example, consulting activities for other firms or the government. This is effort that does not benefit the organization and substitutes for effort exerted to increase the organization’s output. As an example of complementarity, employees can participate in training activities, which provide additional qualifications or skills that are privately beneficial, but may also reduce the marginal cost of effort devoted to inside activities.

### 2.1 Information Structure and Equilibrium Concept

The organization aims at maximizing realized output. It cannot however rely on external enforcement to legally bind members to cooperate. Therefore, the only way to enforce cooperation is through repeated interactions. We follow the literature on self-enforcing agreements by casting the problem in an infinitely repeated-game setting. Let \( \delta \in (0, 1] \) be the common discount factor and the Public Perfect Equilibrium (hereafter PPE) serve as the equilibrium concept of the game between successive generations. A PPE induces a mapping from public history to the intertemporal utility of each member.\(^\text{14}\)

\(^{13}\)A simple parameterization of the cost is the quadratic form \( C(a, b) = c \left[ a^2 + 2kab + b^2 \right] \) where \( c > 0 \), since this allows for the case of substitutability when \( k > 0 \) and the case of complementarity otherwise.

\(^{14}\)See Fudenberg and Tirole (1991) for a formal definition of PPE.
According to Mailath and Samuelson (2006), we allow for a publicly observable randomization \( \phi_t \in [0, 1] \) drawn from a uniform distribution at the beginning of each period, which serves to convexify the set of individuals’ payoffs. The history of realizations of the organization’s output, self-commitment actions, and public randomization up to time \( t \) are respectively denoted by \( h^t_g := (g_0, g_1, ..., g_{t-1}) \), \( h^t_b := (b^0_i, b^1_i, ..., b^{t-1}_i)_{i \neq o} \), and \( h^t_\phi := (\phi_0, \phi_1, ..., \phi_t) \). Members’ strategies map each public history \( h^t := (h^t_g, h^t_b, h^t_\phi) \) to the actions space \( \{a, \pi\} \times \mathbb{R}_+ \), where we refer to \( g_s(h^t), b_s(h^t), \) and \( \phi_s(h^t) \) as the realizations \( g_s, b_s, \) and \( \phi_s \) in the history \( h^t \) at time \( s \leq t \) with \( s \geq 0 \). Since individual strategies are contingent on the history of all preceding plays, self-commitment actions involve a strategic choice to manipulate other members’ equilibrium play. We will refer to this regime as organization-based.

We wish to evaluate the strategic role played by self-commitment actions in promoting cooperation. To do so, we contrast the organization-based regime with the individual-based regime. In the latter regime, members simply ignore any signal related to self-commitment actions in the preceding plays. Therefore, \( h^t_b \) is not part of the relevant history. Namely, members’ strategies map each restricted public history \( \tilde{h}^t := (h^t_g, h^t_\phi) \) to the actions space \( \{a, \pi\} \times \mathbb{R}_+ \), where we refer to \( g_s(\tilde{h}^t), b_s(\tilde{h}^t), \) and \( \phi_s(\tilde{h}^t) \) as the realizations \( g_s, b_s, \) and \( \phi_s \) in the history \( \tilde{h}^t \) at time \( s \leq t \) with \( s \geq 0 \). It follows that in the individual-based regime member’s intertemporal utility is independent of the belief over past self-commitment actions. Therefore, self-commitment actions are chosen as individual static best responses.

### 2.2 Benchmarks

Before embarking on the analysis of the strategic role of self-commitment actions, it is worthwhile making a few observations. Absent altruistic motives regarding future generations, the old member is always better off by shirking than cooperating, since she will exit the organization before the next period.

**Proposition 0.** In any PPE, the efforts expended by the old member are \( (a^o_t, b^o_t) = (a, b^{aut}) \) where \( b^{aut} \) solves \( \theta = C_b (a, b^{aut}) \).

The old member does not contribute to the organization’s output, while maximizing her own private benefit. Her presence, however, is of value to the organization since her successor is able to form expectations of future payoffs, which she will benefit from when old. Depending on whether the successor cooperates or shirks, the expected utility of a member when old is equal to \( \omega := \lambda G (\pi + a) + \theta b^{aut} - C (a, b^{aut}) \) or \( \omega := \lambda G (2a) + \theta b^{aut} - C (a, b^{aut}) \), respectively. Using Proposition 0, we can then rewrite the per-period...
utility of the young member simply as a function of her own effort levels, i.e., \( u(a_t^y, b_t^y) := \lambda (G(a_t^y + a) + \epsilon) + \theta b_t^y - C(a_t^y, b_t^y) \). Hereafter, we shall omit the superscript \( y \) since the young member is the only one who might cooperate.

Having characterized the old member’s behavior, we now turn to establishing two outcomes which will serve as benchmarks. First, consider the scenario in which young members act non-cooperatively in each period, i.e., \( a_t = a \). We refer to this scenario as generational autarky. In such a scenario, the self-commitment action expended by the young member is \( b^{\text{aut}} \) and her intertemporal utility is \( v^{\text{aut}} := u(a, b^{\text{aut}}) + \delta \omega \). Second, suppose that the organization can rely on external enforcement. It would then be optimal for the organization to have both the young and old members cooperate, but the organization has no method for inducing the old member to exert effort on the collective behalf. Hence, the best outcome that the organization can hope for is to have the young member cooperate, i.e., \( a_t = \bar{a} \). We refer to this scenario as constrained first best. In such a scenario, the self-commitment action expended by the young member is \( b^* \), which solves \( \theta = C_b(\bar{a}, b^*) \), and her intertemporal utility is \( v^* := u(\bar{a}, b^*) + \delta \omega \).

The following lemma shows that the level of self-commitment actions expended by the young member in generational autarky is lower than that in the constrained first best in the case of complementarity. The opposite holds true in the case of substitutability.

**Lemma 1.** If \( \Delta(b) < (>) 0 \), then \( b^* > (<) b^{\text{aut}} \).

**Proof.** (See Appendix).

To create scope for cooperation, we assume that intertemporal utility is larger in the constrained first best than in generational autarky, i.e., \( v^* > v^{\text{aut}} \), although the young member’s per-period utility is higher when members shirk rather than cooperate, i.e., \( u(a, b^{\text{aut}}) > u(a, b^*) \). This implies that \( \delta \in (\delta, 1] \) with \( \delta := (u(a, b^{\text{aut}}) - u(a, b^*)) / (\bar{\omega} - \omega) \). As we will see, however, the restriction on the common discount factor does not guarantee that cooperation is sustained in equilibrium.

### 3 Self-Enforcing Intergenerational Cooperation

We now turn to considering the case in which no external enforcement is available. We note that strategies contingent on past histories and enforced by reputation allow for multiple equilibria. However, we confine our attention to the characterization of the best PPE, i.e., the one that simultaneously yields the upper bound of individual intertemporal utility and maximizes the organization’s expected output. Depending on whether or not strategies are contingent on the history of self-commitment actions, two different upper
bounds emerge. In order to determine what they are, we need to first characterize the worst PPE, which is common to both the individual- and organization-based regimes.

3.1 The Worst Equilibrium

The worst PPE provides members with the lowest intertemporal utility. This utility level coincides with $v^{aut}$ according to the following argument: First, $v^{aut}$ is sustainable. Indeed, if it is known that no one will ever cooperate, it is individually optimal not to cooperate. Second, there can be no lower equilibrium payoff, since each member is at the reservation utility.

Proposition 1. The worst PPE is $v^{aut}$, which always exists.

We will use the worst equilibrium as a threat to enforce better equilibria by attaching punishments to signals that are especially likely to arise in the event of a deviation. While deviations from the perfectly observable action $b^t$ must be punished with certainty, deviations from the imperfectly observable action $a^t$ must be deterred with some probability. Randomization of punishment under imperfect monitoring is required to minimize inefficient punishments, which occur on the equilibrium path. Hence, we can with no loss of generality focus on such a trigger strategy to determine the upper bound of PPE intertemporal utilities.\textsuperscript{15}

3.2 The Best Equilibrium in the Individual-Based Regime

We start the analysis of the upper bound of PPE intertemporal utility, denoted by $\tilde{v}^e$, for the case in which members’ strategies are contingent on the history $\tilde{h}^t$. This implies that, regardless of past history, the equilibrium self-commitment action is at the level that maximizes individual per-period utility, namely $b^*$ when members cooperate and $b^{aut}$ when members shirk.

Provided that $\tilde{v}^e \geq v^{aut}$, the best PPE is achieved by a simple trigger strategy, which can be represented by automata with two states: a cooperation state and an absorbing punishment state. Members start in the cooperation state. Regardless of the history $h_b^t$, the punishment state is activated with probability $1 - \phi$ when the organization’s output in the previous periods is lower than a threshold level $\hat{g}$. Namely, at $t = 0$:

(1) Start in the cooperation state by playing $\bar{a}$ and $b^*$;

\textsuperscript{15}The equilibrium strategy is along the lines of Abreu, Pearce, and Stacchetti (1986) who characterize the optimal symmetric equilibrium under imperfect monitoring and binary actions. For a review, see Mailath and Samuelson (2006, pp. 239).
At each $s \leq t$:

(2) If the organization’s output is high, i.e., $g_s \left( \hat{h}^t \right) > \hat{g}$, then go back to (1);

(3) If the organization’s output is low, i.e., $g_s \left( \hat{h}^t \right) \leq \hat{g}$, then go back to (1) with probability $\phi$ or revert permanently to generational autarky with probability $1 - \phi$.

Since old members never cooperate, a low output is realized with probability $p := F ( \hat{g} - G (2a))$ when the young shirk and $q := F ( \hat{g} - G (\bar{a} + a))$ when the young cooperate. Note that the monotone likelihood ratio property holds, i.e., $p > q$, and therefore the likelihood ratio is $L := p/q > 1$.

Calculating the best PPE is equivalent to finding the largest $\phi$ for which the intertemporal utility of each young member is maximized and the self-enforcement constraint is satisfied. Formally:

$$\tilde{\phi}^* := \arg \max_{\phi} \tilde{v} (\phi), \quad (P1)$$

where

$$\tilde{v} (\phi) := \max_{b_t} u (\overline{a}, b_t) + \delta [\overline{\omega} - q (1 - \phi) (\overline{\omega} - \omega)], \quad (1)$$

subject to the constraint:

$$\tilde{v} (\phi) \geq \max_{b_t} u (\underline{a}, b_t) + \delta [\overline{\omega} - p (1 - \phi) (\overline{\omega} - \omega)]. \quad (2)$$

Eq. (1) defines the intertemporal utility in the cooperation state. Inequality (2) describes the self-enforcement constraint regarding the cooperative action, which must be satisfied in order to discourage deviation from $\overline{a}$. Its right-hand side captures the intertemporal utility when agents deviate by choosing $\underline{a}$.\footnote{In this setting, the sufficient condition for the existence of a unique $\hat{g}$ is guaranteed by the monotone likelihood ratio property under which a tail test is the optimal statistical criterion for players to adopt (Abreu, Pearce and Stacchetti, 1990). We could endogenize the level $\hat{g}$ by maximizing $L$ subject to the self-enforcement constraint. Optimal monitoring possibilities, however, are not accessible in practice due to technological constraints. In the following, we treat $p$ and $q$ as exogenous and perform comparative statics with respect to $L$.}

Manipulating Eqs. (1) and (2) yields:

$$\delta (1 - \phi) (p - q) (\overline{\omega} - \omega) \geq u (\underline{a}, b^{aut}) - u (\overline{a}, b^*) > 0. \quad (3)$$

The left-hand side of (3) is the discounted value of avoiding a breakdown in cooperation for each member. It is decreasing in $\phi$ and goes to zero when $\phi$ approaches one. This condition is trivially satisfied since no cooperation is the dominant strategy.\footnote{In equilibrium, it must also be true that members prefer to play $\underline{a}$ rather than $\overline{a}$ in the punishment state. However, this condition is trivially satisfied since no cooperation is the dominant strategy.}
implies that cooperation cannot be sustained without punishment.\(^{18}\) A lower \(\phi\), however, reduces members’ intertemporal utility, as well as the organization’s expected output since cooperation is resumed with a lower probability. This defines the tradeoff between *efficiency* and *enforceability*, which must be resolved.

**Proposition 2.a. (Necessary Condition)** Assume a PPE exists in which \(\tilde{v}^e \geq v^{\text{aut}}\) and \(a_t = \bar{a} \forall t\). Then, \(\phi \in [0, \bar{\phi}]\) with \(\bar{\phi} \geq 0\) and equal to

\[
\bar{\phi} := 1 - \frac{u(a, b^{\text{aut}}) - u(\bar{a}, b^*)}{\delta (p - q) (\bar{\omega} - \omega)}.
\] (4)

In this case, the best PPE is unique and characterized by \(\bar{\phi} = \bar{\phi}\) and

\[
\tilde{v}^e = u(\bar{a}, b^*) + \delta \bar{\omega} - \frac{u(a, b^{\text{aut}}) - u(\bar{a}, b^*)}{L - 1}.
\] (5)

If \(\bar{\phi} < 0\), then the best such PPE yields \(\tilde{v}^e = v^{\text{aut}}\).

**Proof.** (See Appendix).

The best PPE is attained when \(\bar{\phi} = \bar{\phi}\). In fact, there can be no \(\bar{\phi} < \bar{\phi}\), since otherwise Eq. (1) might be further increased without violating constraint (2). Like Abreu, Pearce, and Milgrom (1991), the upper bound (5) is equal to \(v^*\) minus the efficiency loss associated with the inefficient punishment, which occurs with some probability along the equilibrium path. For \(L\) approaching infinity, the efficiency loss associated with the inefficient punishment vanishes and \(\tilde{v}^e\) tends to \(v^*\).

It now remains to characterize the conditions under which \(\bar{\phi}\) is feasible. Note that \(\bar{\phi} \leq 1\) since \(u(a, b^{\text{aut}}) \geq u(\bar{a}, b^*)\). Then, the unique condition to be verified is \(\bar{\phi} \geq 0\). Using Eq. (4), the nonnegativity condition is satisfied when:

\[
\delta \geq \tilde{\delta}^a := \frac{u(a, b^{\text{aut}}) - u(\bar{a}, b^*)}{(p - q) (\bar{\omega} - \omega)}.
\] (6)

If \(\delta < \tilde{\delta}^a\), then \(\bar{\phi} < 0\) and the unique sustainable equilibrium is generational autarky.

**Proposition 2.b. (Sufficient Condition)** A threshold level \(\tilde{\delta}^a \in [\delta, 1]\) exists, so that \(\tilde{v}^e \geq v^{\text{aut}}\) can be sustained as a PPE for any \(\delta \geq \tilde{\delta}^a\).

**Proof.** (See Appendix).

\(^{18}\)Similar implications hold when \(\delta\) approaches zero or when \(p\) tends to \(q\). Indeed, in both cases, the temptation for an agent to deviate is so strong—either because of the agents’ shortsightedness or because of the presence of less informative signals—that no degree of punishment has any deterrent power.
3.3 The Best Equilibrium in the Organization-Based Regime

We now turn to characterizing the upper bound of PPE intertemporal utility, denoted by $v^e$, when members’ strategies are contingent on the history $h^t$. In this scenario, agents choose self-commitment actions while internalizing the strategic impact of such decisions on future members’ equilibrium play.

Provided that $v^e \geq v^{aut}$, the best PPE is achieved by a trigger strategy that shares similar features with the strategy of Section 3.2. However, punishment can now be triggered in the event of a deviation from either cooperation or the agreed-upon level of self-commitment action $b$. This implies that the corresponding automata representation is characterized by three states: a cooperation state and two absorbing punishment states, one for each type of deviation. Namely, at $t = 0$:

(1) Start in the cooperation state by playing $\bar{a}$ and $b$;

At each $s \leq t$:

(2) If the organization’s output is high, i.e., $g_s(h^t) > \hat{g}$, and $b_s(h^t) = b$, then go back to (1);

(3i) If the organization’s output is low, i.e., $g_s(h^t) \leq \hat{g}$, and $b_s(h^t) = b$, then go back to (1) with probability $\phi$ or revert permanently to generational autarky with probability $1 - \phi$;

(3ii) If $b_s(h^t) \neq b$ for any $g_s(h^t)$, then revert permanently to generational autarky with certainty.

The best PPE is achieved here by choosing the appropriate level of $b$ and $\phi$ such that the intertemporal utility of each young member is maximized and the self-enforcement constraints are satisfied. Formally:

$$(b^e, \phi^e) := \arg \max_{(b,\phi)} v(b, \phi),$$  \hspace{1cm} (P2)

where

$$v(b, \phi) := u(\bar{a}, b) + \delta \left[ \omega - q (1 - \phi) (\omega - \omega) \right],$$  \hspace{1cm} (7)

subject to the constraints:

$$v(b, \phi) \geq u(a, b) + \delta \left[ \omega - p (1 - \phi) (\omega - \omega) \right],$$  \hspace{1cm} (8)

$$v(b, \phi) \geq v^{aut}. \hspace{1cm} (9)$$
Eq. (7) defines the intertemporal utility in the cooperation state. In equilibrium, it must be that members prefer to cooperate as well as comply with \( b \). Thus, two self-enforcement constraints need not to be violated: The first is inequality (8). Like Eq. (2), it must be satisfied in order to deter deviation from \( a \). Its right-hand side captures the intertemporal utility when agents deviate by choosing \( a \), albeit complying with the action \( b \). The second is inequality (9), which is the distinctive feature of (P2) and describes the self-enforcement constraint regarding the self-commitment action. It must be satisfied in order to ensure compliance with \( b \). In the case of deviation, punishment is triggered with certainty, given the observability of the action. The individual’s response to certain punishment is then to deviate from cooperation as well. Such an inequality acts as a participation constraint, which, if not satisfied, implies that agents have no incentive to join the organization.

Combining Eqs. (7) and (8) yields the following inequality:

\[
\delta (1 - \phi) (p - q) (\omega - \omega) \geq u(a, b) - u(\bar{a}, b) > 0. \tag{10}
\]

The key difference with inequality (3) is the right-hand side. If \((u(a, b) - u(\bar{a}, b)) - (u(a, b^{aut}) - u(\bar{a}, b'))\) is larger than zero, then the tradeoff between efficiency and enforceability is even exacerbated in this scenario as compared to the individual-based regime. Indeed, cooperation is enforced only if \( \phi < \tilde{\phi} \), with the effect of depressing the individual’s intertemporal utility. The reverse holds true if the difference is negative.

**Proposition 3.a. (Necessary Condition)** Assume a PPE exists in which \( v^e \geq v^{aut} \) and \( a_t = \bar{a} \forall t \). Then, \( \phi \in [\phi(b), \bar{\phi}(b)] \neq \emptyset \) for some \( b \) with

\[
\bar{\phi}(b) := 1 - \frac{u(a, b) - u(\bar{a}, b)}{\delta (p - q) (\omega - \omega)} \tag{11}
\]

and

\[
\phi(b) := \frac{u(a, b^{aut}) - u(\bar{a}, b)}{\delta q (\omega - \omega)} - \frac{1 - q}{q}. \tag{12}
\]

In this case, the best PPE is unique and characterized by \( \phi^e = \tilde{\phi}(b^e) \) where \( b^e \) solves \( u_\phi(\bar{a}, b^e) = \Delta(b^e) / (L - 1) \) and

\[
v^e = u(\bar{a}, b^e) + \delta \omega - \frac{u(a, b^e) - u(\bar{a}, b^e)}{L - 1}. \tag{13}
\]

If \([\phi(b), \bar{\phi}(b)] = \emptyset \forall b\), then the best such PPE yields \( v^e = v^{aut} \).

**Proof.** (See Appendix).

\[19\]For the strategy to be an equilibrium, it must also be true that in the punishment states agents prefer to play \( a \) rather than \( \bar{a} \), but this is trivially satisfied since \( u(a, b) > u(\bar{a}, b) \) for any \( b \).
Proposition 3.a predicts under which conditions \( v^e \geq v^\text{aut} \). Suppose that a nonempty set of pairs \((b, \phi)\) that satisfy constraints (8) and (9) exists. Then, in equilibrium the following facts are necessarily true: There can be no \( \phi > \phi(b) \); otherwise, inequality (8) would be violated. This would occur because of the temptation not to cooperate. Analogously, there can be no \( \phi < \phi(b) \); otherwise, inequality (9) would be violated. This would occur because of the temptation not to join the organization. From Eq. (7), moreover, we learned that intertemporal utility is increasing in \( \phi \). We then conclude that the best PPE is attained for \( \phi = \phi(b) \). Conditional on \( \phi(b) \) being feasible, the following corollary holds:

**Corollary 1.** When privately and collectively beneficial efforts are complements (substitutes):

(i) \( \phi(b) \) is strictly increasing (decreasing) in \( b \);

(ii) For a given \( b \), \( \phi(b) \) is larger when \( \delta \) and \( p \) are larger and \( q \) is smaller.

**Proof.** (See Appendix).

Part (i) of Corollary 1 is obtained by differentiating \( \phi(b) \) with respect to \( b \), which yields \( \phi_b = -\Delta(b) / (\delta(p-q)(\omega-\omega)) \). Therefore, \( \phi_b > 0 \) when \( \Delta(b) < 0 \) and \( \phi_b < 0 \) otherwise. In the case of complementarity, an increase in \( b \) leads to more forgiveness and in turn to a higher expected utility when old. Indeed, once agents have internalized that a larger \( b \) reduces the individual gain from abandoning cooperation and have observed that previous members have distorted up the level of the self-commitment action, they consistently believe that the realization of a low output is to be attributed primarily to a negative shock. The reverse argument applies when efforts are substitutes. Part (ii) of Corollary 1 states that \( \phi(b) \) is lower and, in turn, cooperation is harder to sustain when agents are less patient, i.e., \( \delta \) is smaller, or monitoring is weaker, i.e., \( p \) approaches \( q \). In these cases, members are more tempted to cheat. For any given \( b \), generations must therefore punish with higher probability in order to discourage deviation from cooperation.

**Corollary 2.** When privately and collectively beneficial efforts are complements (substitutes):

(i) \( b^e > (\leq) b^* \);

(ii) \(|b^* - b^e| \) decreases with \( L \).

**Proof.** (See Appendix).

Corollary 2 provides two additional insights: Part (i) implies that \( b^e \) always differs from the constrained first-best level, regardless of the type of interaction between efforts.
Figure 1: The best PPE for $\Delta (b) > 0$ (left-side panel) and $\Delta (b) < 0$ (right-side panel).

This occurs because a distorted level of self-commitment action improves the continuation utility by reducing the inefficiency loss associated with the punishment, i.e., it enlarges $\bar{\phi} (b)$, even though it reduces the young member’s per-period utility in the cooperation state, i.e., it lowers $u (a, b)$. Part (ii) predicts that a member’s motive to signal her willingness to cooperate by distorting self-interested actions vanishes when the organization’s output is a good indicator of individual contributions.

Figure 1 illustrates these results. The shaded area depicts the set of pairs $(b, \phi)$ that satisfy constraints (8) and (9), whose upper boundary is downward-sloping when efforts are substitutes and upward-sloping otherwise. The graph also plots the iso-utility curve of (7). The intertemporal utility increases as $\phi$ increases and $|b^* - b^e|$ shrinks. Therefore, it is maximized at point $A$. However, such a maximum is not enforceable since constraint (8) is violated at this point. The best PPE is characterized for a given $L$ by the tangency point between the iso-utility curve and $\bar{\phi} (b)$, namely, at point $E$. At such a point, $b^r \neq b^*$ and is located to the left of it when $\Delta (b) > 0$ and to the right of it otherwise. Finally, the graph depicts the locus $(b^r, \phi^r)$ resulting from variations in $L$ and represented by the dotted red line, which approaches point $A$ as $L$ tends to infinity.

It now remains to determine when $[\bar{\phi} (b^r), \bar{\phi} (b^r)] \neq \emptyset$. This is guaranteed whenever the conditions $\bar{\phi} (b^r) \geq 0$ and $\bar{\phi} (b^r) \geq \bar{\phi} (b^r)$ simultaneously hold.\textsuperscript{20} Using Eqs. (11) and (12), this is the case when:

$$\delta \geq \delta^a := \frac{u (a, b^r) - u (\bar{a}, b^r)}{(p - q) (Z - \omega)}$$  \hspace{1cm} (14)

\textsuperscript{20}Note that $\bar{\phi} (b^r) < 1$. Moreover, if $\bar{\phi} (b^r) \geq \bar{\phi} (b^r)$, then $\bar{\phi} (b^r) \leq 1$. 

17
and
\[
\delta \geq \delta^b := \frac{u(a, b^{\text{aut}}) - u(\bar{\pi}, b^e)}{\bar{\omega} - \bar{\omega}} + \frac{q}{p - q} \frac{u(a, b^e) - u(\bar{\pi}, b^e)}{\bar{\omega} - \bar{\omega}}.
\] (15)

If \( \delta < \delta^a \), then \( \phi(b^e) < 0 \). This implies that constraint (8) is violated and therefore \( v^c = v^{\text{aut}} \). Similarly, if \( \delta < \delta^b \), then \( \phi(b^e) < \phi(b^e) \). This implies that members have no incentive to join the organization when a level of self-commitment action equal to \( b^e \) is required. Once again, the best PPE intertemporal utility is \( v^{\text{aut}} \). Depending on \( \max\{\delta^a, \delta^b\} \), either constraint (8) or (9) can bind first. In both cases, we can establish the following result:

**Proposition 3.b. (Sufficient Condition)** A threshold level \( \max\{\delta^a, \delta^b\} \in [\delta, 1] \) exists, so that \( v^c \geq v^{\text{aut}} \) can be sustained as a PPE for any \( \delta \geq \max\{\delta^a, \delta^b\} \).

**Proof.** (See Appendix).

4 Strategic Value of Self-Commitment Actions

We have so far determined the best PPE for each of the two regimes. While forgiveness is linked to the level of self-commitment action in the organization-based regime, it is immune to strategic manipulation in the individual-based regime. The following question then naturally arises: Which of the two regimes provides the most value to the organization as well as to its members?

**Proposition 4.** Strategic self-commitment actions are always of value to the organization since \( \bar{\phi}(b^e) < \phi(b^e) \).

**Proof.** (See Appendix).

In equilibrium, requiring members to build up their reputations by means of outside tasks minimizes their gains by deviating from cooperation. The smaller are the members’ gains from shirking, the higher will be the probability that agents forgive, as described in Eq. (11). More forgiveness implies that cooperation is more likely in the long run and that the organization’s expected output is larger. For this reason, motivating members to pursue self-interested activities in order to signal their willingness to cooperate is the best practice for the organization to secure long-term engagement.

Unlike the organization, members face the following intertemporal tradeoff associated with being in the organization-based regime rather than the individual-based regime: a short-term loss due to the reduction in per-period utility when young, as implied by Corollary 2, versus a long-term gain associated with a higher expected utility when old, as derived from Proposition 4. To evaluate which of the two effects prevails, we introduce
the members’ surplus function \( W : [\delta, 1] \to \mathbb{R} \), defined as the difference between the payoffs (5) and (13), i.e., \( W := v^c - \tilde{v}^c \). The following proposition shows that exerting strategic self-commitment actions increases members’ well-being, i.e., increases efficiency (Part (i)), and at the same time expands the scope of cooperation, i.e., improves enforceability (Part (ii)).

**Proposition 5.** Strategic self-commitment actions are always of value to the organization’s members, i.e.:

1. \( W > 0 \) for any \( \delta > \max\{\tilde{\delta}^a, \max\{\delta^a, \delta^b\}\} \);
2. \( \tilde{\delta}^a > \max\{\delta^a, \delta^b\} \).

**Proof.** (See Appendix).

The result of Part (i) is intuitive. Since the individual’s incentive to abandon cooperation relies on the gain from defecting, reducing such a gain by means of self-commitment actions is the most efficient way to relax the binding self-enforcement constraint on the cooperative action. Hence, members, when choosing self-commitment actions so as to maximize their intertemporal utility, fully internalize the strategic impact of self-interested activities on equilibrium forgiveness and future cooperative incentives. As a consequence, the resulting intertemporal utility is larger than the level attained by any other value of \( b \). In particular, it turns out to be larger than the intertemporal utility attained if self-commitment actions are taken as static best responses in both cooperation and punishment states. This result holds for any degree of monitoring imperfection.

Part (ii) of Proposition 5 predicts that when agents are so impatient that cooperation cannot be sustained, i.e., \( \delta < \tilde{\delta}^a \), strategies contingent on the history of self-commitment actions can improve enforceability. This is because in the organization-based regime cooperation is self-enforced when \( \delta \geq \max\{\delta^a, \delta^b\} \), where \( \max\{\delta^a, \delta^b\} \) proves to be smaller than \( \tilde{\delta}^a \). It turns out that the members’ surplus is positive also for \( \delta \in [\max\{\delta^a, \delta^b\}, \tilde{\delta}^a) \), whereas it is equal to zero for \( \delta < \max\{\delta^a, \delta^b\} \).

Figure 2 illustrates the results of Proposition 5. The upper graphs plot the upper bounds of intertemporal utilities in each of the two regimes. The solid-black line denotes \( v^c \), while the dashed-red line denotes \( \tilde{v}^c \). Both upper bounds are equal to \( v^{aut} \) insofar as \( \delta \in (\tilde{\delta}, \max\{\delta^a, \delta^b\}) \) in the case of \( v^c \) and \( \delta \in (\tilde{\delta}, \tilde{\delta}^a) \) in the case of \( \tilde{v}^c \). After that, they may jump to a higher value and monotonically converge to their maximum sustainable values when \( \delta \) approaches one.\(^{21}\) The lower graphs plot the members’ surplus \( W \).

\(^{21}\)A formal discussion of the discontinuity in the map \( W \) is provided in the proof of Proposition 2.b and 3.b in the Appendix.
4.1 Mandatory Provision

The previous analysis showed that the strategic interaction between collectively and privately beneficial efforts is of value to the organization, as well as to its members, and this is true whether the two efforts are complements or substitutes. A question that then arises is whether the introduction of a mandatory provision constraint on the cooperative action can further improve the organization’s performance and the members’ well-being. These types of constraints are common in organizations. For example, workers in a firm can be required to carry out a minimum number of tasks or to work a minimum number of hours.

To introduce mandatory provision, we slightly modify the basic model by considering a compulsory minimum collectively beneficial effort $a_{\min} > a$. The larger $a_{\min}$, the tighter is the mandatory provision constraint. The imposition of such a constraint clearly increases the organization’s expected output, since it reduces the individual’s gain of deviation from cooperative behavior. However, it has a negative effect on the members’ surplus, as stated in the following proposition:

**Proposition 6.** When $a_{\min}$ is increased:

(i) $|b^e - b^*|$ shrinks;
(ii) $W$ is reduced.

**Proof.** *(See Appendix).*

The result of Proposition 6 has interesting implications in terms of how formal institutions should be modelled in the presence of informal rules within an organization. Indeed, the introduction of a minimum provision constraint undermines rather than promotes the self-enforcement of cooperation when self-interested actions involve a strategic decision to manipulate future cooperative incentives. Although the introduction of a mandatory provision constraint forces members to contribute more in the case of low cooperation and in turn decreases the short-term gain from opportunistic behavior, it lowers the marginal benefit associated with the higher level of forgiveness implied by compliance with a self-commitment action. This in turn negatively affects the members’ surplus.\(^{22}\)

## 5 Multiple Generations

The baseline model rests on the assumption that only two generations inhabit the organization at any point in time. We now extend it to include more than two generations. The model can then be applied to the description of the optimal task assignment along career paths in an organization. We show that it may be optimal to require generations who are more reluctant to cooperate to underperform or overperform in outside tasks so as to motivate them to try harder in tasks that are profitable for the organization.

Consider an organization whose members live for three periods: young, middle-aged, and old. Let $m$ denote the middle-aged member. As before, we focus on the characterization of the best PPE, which requires the largest number of generations to choose $π$ and yields the upper bound of individual intertemporal utility. This can be accomplished by using the trigger strategy described in Section 3 and appropriately choosing a level of $φ$. Given that individual contributions are unobservable, the identity of the deviating member cannot be inferred by observing the organization’s realized output. Hence, punishment in equilibrium must be triggered with the same probability in all generations.\(^{23}\) Clearly, Proposition 0 still holds and the expected utility of a member when old is equal to $ω := \lambda G (2π + a) + θ b^{\text{aut}} - C (a, b^{\text{aut}})$ or $ω := \lambda G (3a) + θ b^{\text{aut}} - C (a, b^{\text{aut}})$.

\(^{22}\)Bernheim and Whinston (1998) obtain results with a similar flavor for non-intergenerational contracting problems. When complete contracts are impossible and voluntary cooperation is needed in the dimensions where the contract is incomplete, it might be optimal to leave some dimensions out of the contract in order to increase the incentive to cooperate in the dimensions that cannot be included. Rangel (2003) shows that in an overlapping generations framework minimum provision constraints undermine cooperation if it is already sustained by means of informal institutions.

\(^{23}\)The probabilities of generating a bad signal in the case of individual deviation generalizes to $p := F (\bar{g} - G (3a))$ and in the case of cooperation to $q := F (\bar{g} - G (2π + a))$. 
depending on whether other members cooperate or shirk. For notational purposes, let \( u(a^i, b^i) := \lambda (G(a^i + \bar{a} + a) + \epsilon) + \theta b^i - C(a^i, b^i) \) be the per-period utility of member \( i \) when the old member shirks and the other members cooperate. The intertemporal utility in the cooperation state for the middle-aged and young members can therefore be written respectively as:

\[
v^m(\phi, b^m) := u(\bar{a}, b^m) + \delta (\overline{\omega} - q (1 - \phi) (\overline{\omega} - \omega)) \tag{16}
\]

and

\[
v^y(\phi, b^y, b^m) := u(\bar{a}, b^y) + \delta (v^m(\phi, b^m) - q (1 - \phi) (v^m(\phi, b^m) - (u^{aut} + \delta \bar{\omega}))) \tag{17}
\]

where \( u^{aut} := \lambda (G(3a) + \epsilon) + \theta b^{aut} - C(a, b^{aut}) \) is the per-period utility in generational autarky. For the candidate strategy to be an equilibrium, it must be that middle-aged members prefer \( \bar{a} \) and \( b^m \) over deviating to \( a \) and \( b^{aut} \). Formally,

\[
v^m(\phi, b^m) \geq u(a, b^m) + \delta (\overline{\omega} - p (1 - \phi) (\overline{\omega} - \omega)) \tag{18}
\]

and

\[
v^m(\phi, b^m) \geq u^{aut} + \delta \bar{\omega}.
\]

Likewise, it must be the case that in the cooperation state young members prefer \( \bar{a} \) and \( b^y \) over deviating to \( a \) and \( b^{aut} \), i.e.:

\[
v^y(\phi, b^y, b^m) \geq u(a, b^y) + \delta (v^m(\phi, b^m) - p (1 - \phi) (v^m(\phi, b^m) - (u^{aut} + \delta \bar{\omega}))) \tag{19}
\]

and

\[
v^y(\phi, b^y, b^m) \geq u^{aut} + \delta (u^{aut} + \delta \bar{\omega}).
\]

For simplicity of exposition, we focus on the self-enforcement constraints on the cooperative decision for young and middle-aged members.\(^{24}\) Since Eq. (17) is increasing in \( \phi \), the best PPE involves the selection of the maximum level of forgiveness satisfying constraints (18) and (19). We note that an increase in \( \phi \) has a twofold impact on the cooperative behaviour of members. On the one hand, it lowers the probability of cooperation breakdown, thereby making deviation more profitable for both generations. On the other hand, it raises \( v^m(\phi, b^m) \) thereby reducing the temptation of the young member to

\(^{24}\)In the previous sections, we saw that depending on parameters, either the self-enforcement constraint on the cooperative decision or the self-enforcement constraint on the self-commitment action can bind first. In this section, we confine our attention to the case in which compliance with a self-commitment action is necessarily guaranteed insofar as compliance with cooperation is enforced.
shirk. Which of the two effects prevails depends on $\delta$ and identifies the generation that is the most reluctant to cooperate. Let $b^{i,e}$ denote the equilibrium level of self-commitment action for member $i \in \{y, m, o\}$. Then, the following result holds:

**Proposition 7.** The best PPE satisfying $a^i = \bar{a}$ for each $i$ is characterized by:

(i) $b^{y,e} = b^*$, $b^{m,e} > ( < ) b^*$ when $\Delta (b^m) < ( > ) 0$, and $b^{o,e} = b^{aut}$ when the middle-aged member is the most reluctant;

(ii) $b^{y,e} > ( < ) b^*$ when $\Delta (b^y) < ( > ) 0$, $b^{m,e} = b^*$, and $b^{o,e} = b^{aut}$ when the young member is the most reluctant.

**Proof.** (See Appendix).

Proposition 7 suggests that organizations seeking to improve workplace engagement of all employees should assign different tasks to members at different stages of their career. When the discount factor is sufficiently large, the junior members always comply with cooperation whenever the mid-career members also do. It follows that members in the early stage of their careers are highly motivated to build long-term engagement in the organization. In this scenario, incentives provided by means of optimal task assignment in outside activities should be designed so as to motivate more experienced members to exert high effort in tasks that are of value to the organization. Part (i) predicts that the lifetime profile of self-commitment actions is either hump-shaped when privately and collectively beneficial efforts are complements or inverted hump-shaped when they are substitutes. In contrast, when the discount factor is overly small so that agents have low expectations of eventually being promoted to most senior positions, it is optimal for the organization to allow members to pursue outside tasks and thus build up their reputations early in their career in order to boost their motivation to comply with cooperation. As stated in Part (ii), the resulting lifetime profile of self-commitment actions is then either monotonically decreasing or monotonically increasing depending on whether the two efforts are complements or substitutes.

6 Conclusion

The model attempts to capture the role played by self-interested actions in promoting cooperation in an organization when individual contributions to the common good are imperfectly observable. Two fundamental features of the model drive the results: the interaction among multiple activities and the repeated interaction among members. Agents
distort self-interested actions only if they anticipate that such a choice will positively affect future cooperative incentives and thus raise their continuation utility.

These two features fundamentally distinguish the model from existing theories of commitment. Commitment devices may improve cooperation in a static setting without requiring interdependent actions. However, the potential gains from cooperation are unlikely to be realized for the following three reasons: First, maintaining cooperation may require mutual pledges of commitment through written contracts, which are often difficult to enforce. Second, commitment is profitable only if it is not overly costly and not activated along the equilibrium path. With monitoring imperfection, however, agents must periodically incur the cost that they have committed to. Finally, the commitment solution requires the existence of an outside entity to construct commitment spaces for the players, such that cooperation is attainable as an equilibrium outcome in the restricted action space. In contrast to existing theories of commitment, our mechanism accomplishes cooperation in a plausible fashion. First, self-commitment actions are self-enforcing. Second, they are not conceived as a punishment technology, but rather as an equilibrium strategy, which we show to always have value to both an organization and its members in the presence of imperfect monitoring. Finally, self-commitment actions render unnecessary the existence of an outside entity who is able to construct commitment spaces since they are part of the players’ strategy space.

The idea of self-commitment actions was used to study the sustainability of cooperation in ongoing organizations, such as public agencies and private firms. Clearly, there are many other interesting applications that fit the setting. An example is the study of socially responsible practices in corporations, which can be viewed as an optimal communal response in an uncertain environment, such as a volatile financial market. If such socially responsible practices reduce gains from unobservable deviation, then they can also foster goodwill and trust among shareholders (see, e.g., Baron, 2001 and 2010). Another prominent example might be the study of cooperation in religious organizations, which are ongoing and exist in order to achieve the common good of their members (see, e.g., Iannaccone, 1992; Levy and Razin, 2012). In this context, the model might be useful in highlighting the role of religious practices which, like self-commitment actions, can dictate members’ behavior.
7 Appendix

Proof of Lemma 1. First consider the case of complementary efforts, i.e., \(C_b(\overline{a}, b^*_i) - C_b(\overline{a}, b^*_j) < 0\). If \(b^*_i = b^*\), then \(\theta = C_b(\overline{a}, b^*) < C_b(\overline{a}, b^*)\) given the convexity of \(C(\cdot)\). It follows that \(b^\text{aut}\) cannot be larger than \(b^*\), otherwise \(\theta < C_b(\overline{a}, b^*) < C_b(\overline{a}, b^\text{aut})\) which would contradict \(\theta = C_b(\overline{a}, b^\text{aut})\). An analogous argument holds true for the case of substitute efforts.

Proof of Proposition 2.a. For a fixed \(\hat{g}\), there are two signals available. Hence, it is not possible to reward one member when the other is punished. It is then optimal to raise forgiveness to the point at which agents are indifferent between complying and shirking. Formally, using Eq. (1), we obtain that condition (2) holds when \(\phi \leq \overline{\phi}\) with \(\overline{\phi}\) as reported in Eq. (4). If a PPE exists in which \(\tilde{\phi}^C \geq v^\text{aut}\), it then must be that \(\tilde{\phi}^C = \overline{\phi}\) with \(\overline{\phi} \geq 0\). Inserting Eq. (4) into (1) yields the upper bound of PPE, i.e., Eq. (5) in the text. If \(\overline{\phi} < 0\), then Eq. (2) is not satisfied and the upper bound is \(\tilde{\phi}^\text{aut}\).

Proof of Proposition 2.b. Since \(p > q\), then \(\tilde{\delta}^a > \hat{\delta}\). Furthermore, \(\tilde{\delta}^a \leq 1\) when

\[
q \leq p - \frac{u(a, b^\text{aut}) - u(\overline{a}, b^*)}{(\overline{\omega} - \omega)}.
\]

For some levels of \(p\) and \(q\) the above condition is satisfied. Hence, there always exists a feasible \(\tilde{\delta}^a\), such that: (i) if \(\delta > \tilde{\delta}^a\), then \(\overline{\phi} > 0\) and \(\tilde{\phi}^C > v^\text{aut}\); (ii) if \(\delta = \tilde{\delta}^a\), then \(\overline{\phi} = 0\) and \(\tilde{\phi}^C > (=) v^\text{aut}\), depending on whether \(p < (=) 1\).

Proof of Proposition 3.a. We use the same logic as in the proof of Proposition 2.a. However, punishment can here be triggered after deviation from either cooperation or self-commitment actions. Given that self-commitment actions are perfectly observable, it is optimal to revert to generational autarky with certainty after a deviation has been observed. Furthermore, it is optimal to exert the level \(b\) that maximizes the individual’s intertemporal utility. Formally, using Eq. (7), we obtain that condition (9) holds when \(\phi \leq \overline{\phi}(b)\), whereas (9) holds when \(\phi \geq \overline{\phi}(b)\), with \(\overline{\phi}(b)\) and \(\phi(b)\) as reported in Eqs. (11) and (12). If a PPE exists in which \(v^C \geq v^\text{aut}\), then it must be that \(\phi = \overline{\phi}(b)\) and \(\overline{\phi}(b) \geq \phi(b)\) for some \(b\). Furthermore, \(b = b^c\) where \(b^c\) solves \(u_b(\overline{a}, b^c) = \Delta(b^c) / (L - 1)\). Inserting Eq. (11) into (7) yields the upper bound of PPE, i.e., Eq. (13) in the text. If \(\overline{\phi}(b) < \phi(b)\) for all \(b\), then there exists no feasible \(\phi\) that simultaneously satisfies Eqs. (8) and (9), and the upper bound is \(v^\text{aut}\).

Proof of Corollary 1. The proof of Part (i) appears in the text. To prove Part (ii), differentiating \(\overline{\phi}(b)\) with respect to \(\delta\), \(p\), and \(q\) yields \(\overline{\phi}_\delta = \frac{u(\overline{a}, b) - u(\overline{a}, b^*)}{\delta(p-q)(\omega - \omega)} > 0\), \(\overline{\phi}_p = \frac{u(a, b^c) - u(\overline{a}, b^*)}{\delta(p-q)(\omega - \omega)} > 0\), and \(\overline{\phi}_q = -\frac{u(a, b^c) - u(\overline{a}, b^*)}{\delta(p-q)^2(\omega - \omega)} < 0\).
Proof of Corollary 2. Start by proving Part (i). The equilibrium level $b^e$ solves $\theta - C_b(\bar{\pi}, b^*) - (1/(L-1)) \Delta(b^e) = 0$. Hence, $\theta - C_b(\bar{\pi}, b^e) > (\leq) 0$ when $\Delta(b^e) > (\leq) 0$. Since $b^*$ solves $\theta - C_b(\bar{\pi}, b^*) = 0$, it follows that $b^e > (\leq) b^*$ given the convexity of $C(\cdot)$. We now prove Part (ii). Using the implicit function theorem yields $b_L^e = -C_b(\bar{\pi}, b^e)/(1/(L-1)\Delta(b^e))$, where the denominator is negative since it is the second-order condition with respect to $b$. Hence, $b_L^e > (\leq) 0$ insofar as $\Delta(b^e) > (\leq) 0$. The result for Part (i) and $b_L^e = 0$ guarantee that $|b^e - b^e|$ decreases with $L$. ■

Proof of Proposition 3.b. For $\phi(b^e) \geq 0$ to be true, condition (14) must hold. Feasibility requires that $\delta^n \leq 1$, i.e.,

$$q \leq p - \frac{u(a, b^e) - u(\bar{\pi}, b^e)}{\omega - \bar{\omega}}. \tag{20}$$

and $\delta^n > \breve{\delta}$, i.e.,

$$q > p - \frac{u(a, b^e) - u(\bar{\pi}, b^e)}{u(a, b^{aut}) - u(\bar{\pi}, b^e)}. \tag{21}$$

Moreover, $\phi(b^e) \geq \phi(b^e)$ implies that condition (15) must hold. In this case, feasibility requires that $\delta^b \leq 1$, i.e.,

$$q \leq \frac{(\omega - \bar{\omega}) - (u(a, b^{aut}) - u(\bar{\pi}, b^e))}{(\omega - \bar{\omega}) - (u(a, b^{aut}) - u(a, b^e))}p. \tag{22}$$

and $\delta^b > \breve{\delta}$, which is always satisfied. A feasible $\delta > \max\{\delta^n, \delta^b\}$ therefore exists if conditions (20), (21), and (22) hold. From Part (i) of Corollaries 1 and 2, we learned that $\phi(b^e) > \phi(b^*)$, which implies $u(a, b^e) - u(\bar{\pi}, b^e) < u(a, b^*) - u(\bar{\pi}, b^*)$. Since $u(a, b^*) < u(a, b^{aut})$, the condition $u(a, b^e) - u(\bar{\pi}, b^e) < u(a, b^{aut}) - u(\bar{\pi}, b^*)$ is true, which guarantees that $\frac{u(a, b^e) - u(\bar{\pi}, b^e)}{u(a, b^{aut}) - u(\bar{\pi}, b^*)} < 1$. Moreover, to guarantee $\delta < 1$ it must be the case that $u(a, b^{aut}) - u(\bar{\pi}, b^*) < \omega - \bar{\omega}$, which implies that $\frac{u(a, b^e) - u(\bar{\pi}, b^e)}{u(a, b^{aut}) - u(\bar{\pi}, b^*)} < \frac{u(a, b^e) - u(\bar{\pi}, b^e)}{\omega - \bar{\omega}}$. For some levels of $p$ and $q$ conditions (20) and (21) are satisfied. Hence, there always exists a feasible max $\{\delta^n, \delta^b\}$, such that if $\delta > \max\{\delta^n, \delta^b\}$, then $\phi(b^e) \geq \phi(b^e)$ and $\phi^e > \phi^{aut}$. Finally, we note that $\delta^n \geq (\leq) \delta^b$ if and only if

$$\frac{u(a, b^{aut}) - u(\bar{\pi}, b^e)}{u(a, b^e) - u(\bar{\pi}, b^e)} \leq (\geq) \frac{1 - q}{p - q}. \tag{23}$$

If $\delta = \delta^n = \max\{\delta^n, \delta^b\}$, then $\phi^e = 0$. In this case, $\phi^e = (\leq) \phi^{aut}$ depending on whether $p < (\geq) 1$. If $\delta = \delta^b = \max\{\delta^n, \delta^b\}$, then $\phi(b^e) = \phi^e = 0$ and $\phi^e = \phi^{aut}$ for any level of $p$. ■

Proof of Proposition 4. From the proof of Proposition 3.b, we learned that $u(a, b^e) -
$u(\pi, b^e) < u(\bar{a}, b^a) - u(\pi, b^*)$. This directly implies that $\phi^e > \tilde{\phi}^e$. ■

**Proof of Proposition 5.** We first prove Part (i). Since $b^e$ solves $u_b(\bar{a}, b^e) = \Delta(b^e) / (L - 1)$, it must be true that $v^e > u(\bar{a}, b^*) + \delta \omega - (1/(L - 1))(u(\bar{a}, b^*) - u(\pi, b^*))$. Furthermore, $u(\bar{a}, b^a) > u(\bar{a}, b^*)$ implies that $u(\bar{a}, b^*) + \delta \omega - (1/(L - 1))(u(\bar{a}, b^*) - u(\pi, b^*)) > \tilde{v}^e$. We then conclude that $W = v^e - \tilde{v}^e$ is positive for any $L$ and $\delta > \max\{\tilde{\delta}^a, \max\{\tilde{\delta}^a, \tilde{\delta}^b\}\}$. We now prove Part (ii). First, let $\delta^a = \max\{\tilde{\delta}^a, \tilde{\delta}^b\}$. Using Proposition 4 and comparing Eq. (6) to Eq. (14), it is straightforward to show that $\tilde{\delta}^a > \delta^a$. Second, let $\tilde{\delta}^b = \max\{\tilde{\delta}^a, \tilde{\delta}^b\}$. If $\tilde{\delta}^a < \tilde{\delta}^b$, then $v^e = v^a$ and $\tilde{v}^e > v^e = v^a$ for $\delta = \tilde{\delta}^b$. By continuity, for any $\epsilon > 0$ such that $\tilde{\delta}^a + \epsilon > \tilde{\delta}^b$, if $\delta = \tilde{\delta}^a + \epsilon$, then $\tilde{v}^e > v^e > v^a$, which contradicts the result of Part (i), namely that $W > 0$ for any $\delta > \max\{\tilde{\delta}^a, \max\{\tilde{\delta}^a, \tilde{\delta}^b\}\}$. ■

**Proof of Proposition 6.** We first prove Part (i). Let $a^\min \in (\bar{a}, \pi]$. Then, the per-period utility is $u(a^\min, b)$ when the young do not cooperate. The equilibrium self-commitment action solves $0 = u_b(\bar{a}, b^e) - (1/(L - 1))(u_b(a^\min, b^e) - u_b(\bar{a}, b^*)))$. The implicit function theorem yields:

$$b^e_{a^\min} = \frac{-1}{u_{bb}(\bar{a}, b^e) - (1/(L - 1))(u_{bb}(a^\min, b^e) - u_{bb}(\bar{a}, b^*)))},$$

where the denominator is the second-order condition, which is satisfied in equilibrium. The sign of $u_{a^\min b}$ is the opposite to that of $\Delta(b^e)$. Since $b^e > (\chi) b^*$ when $\Delta(b^e) < (\chi) 0$ and $b^a_{a^\min} = 0$, we obtain that $|b^e - b^*|$ is decreasing in $a^\min$. We now prove Part (ii). Differentiating $W$ with respect to $a^\min$ and applying the envelope condition yields:

$$W_{a^\min} = \frac{1}{L - 1}\left((u_{a^\min}(a^\min, b^a) - u_{a^\min}(a^\min, b^e)) - u_b(a^\min, b^e) \frac{\partial b^e}{\partial a^\min}\right),$$

$$\simeq - u_{a^\min b}(a^\min, b) [b^e - b^a] - u_b(a^\min, b^e) \frac{\partial b^e}{\partial a^\min}.$$ Using Lemma 1, Part (i) of Corollary 2 and of Proposition 6 yield: (i) $u_{a^\min b} > 0, b^a < b^* < b^e, u_b(a^\min, b^a) < 0$, and $\frac{\partial b^e}{\partial a^\min} < 0$ when $\Delta(b) < 0$; (ii) $u_{a^\min b} < 0, b^e < b^* < b^a$, $u_b(a^\min, b^e) > 0$, and $\frac{\partial b^e}{\partial a^\min} > 0$ when $\Delta(b) > 0$. In both cases, $W_{a^\min} < 0$. ■

**Proof of Proposition 7.** The best PPE is achieved by choosing the appropriate levels of $\phi$, $b^e$, and $b^a$ such that Eq. (17) is maximized and the self-enforcement constraints (18) and (19) are satisfied. We start by proving Part (i). Consider the case in which constraint (18) is binding and constraint (19) is slack. We determine ex-post the conditions under which this case holds true. A binding constraint (18) pins down $\phi$ as a function of $b^a$, i.e.,

$$\phi = \phi^a(b^a) := 1 - \frac{u(a^a, b^a) - u(a, b^a)}{\delta(p - q)(\omega - \omega)}, \quad (23)$$
Note that \( v^\phi (\phi, b^\phi, b^m) \) is increasing in \( v^m (\phi, b^m) \). Then, the equilibrium level \( b^{m,e} \) is chosen so as to maximize (16) with respect to \( b^m \) subject to Eq. (23). Hence, \( b^{m,e} \) solves:

\[
0 = u_{b^m}(\varpi, b^{m,e}) - \frac{q}{p - q} \Delta (b^{m,e}).
\]

The concavity of \( u (\varpi, b^m) \) in \( b^m \) implies that \( b^{m,e} < (>) b^* \) when \( \Delta (b^{m,e}) > (<) 0 \). Finally, the equilibrium level \( b^{m,e} \) is chosen so as to maximize \( v^\phi (\phi, b^\phi, b^m) \) with respect to \( b^\phi \) subject to Eq. (23) and \( b^m = b^{m,e} \). This yields the first-order condition \( 0 = u_{b^\phi}(\varpi, b^{m,e}) \), which implies that \( b^{m,e} = b^* \). We now determine the condition under which constraint (18) is binding and constraint (19) is slack. This is true when:

\[
\delta > \delta^m := \frac{q}{p - q} \frac{u(a, b^{m,e}) - u(\varpi, b^{m,e})}{\varpi - \omega} - \frac{u(\varpi, b^{m,e}) - u^\text{aut}}{\varpi - \omega} + \frac{u(a, b^{\phi,e}) - u(\varpi, b^{\phi,e})}{u(a, b^{m,e}) - u(\varpi, b^{m,e})}.
\]

For some values of \( p \) and \( q \), there exists a feasible \( \delta \) satisfying the above condition. Now consider Part (ii) in which constraint (18) is slack and constraint (19) is binding. A binding constraint (19) pins down \( \phi \) as a function of \( b^\phi \) and \( b^m \), i.e., \( \phi = \phi^\phi (b^\phi, b^m) \). Formally, \( \phi \) is the largest value that solves the following equation:

\[
\phi = 1 - \frac{u(a, b^\phi) - u(\varpi, b^\phi)}{\delta (p - q)(v^m (\phi, b^m) - (u^\text{aut} + \delta \omega))}.
\]

Then, the equilibrium level \( b^{m,e} \) is chosen so as to maximize Eq. (17) with respect to \( b^m \) subject to \( \phi = \phi^\phi (b^\phi, b^m) \). Hence, \( b^{m,e} \) solves:

\[
u_{b^m}(\varpi, b^{m,e}) = - \left( \frac{q}{1 - \phi^\phi (\cdot, b^{m,e})} \left( u^m (\phi^\phi (\cdot, b^{m,e}) - (u^\text{aut} + \delta \omega)) \right) \right) \phi^\phi_{b^m}.
\]

Since the term in the brackets is positive, \( u_{b^m}(\varpi, b^{m,e}) \geq (<) 0 \) if and only if \( \phi^\phi_{b^m} \leq (>) 0 \). By the implicit function theorem, Eq. (24) defines:

\[
\phi^\phi_{b^m} = \frac{(u(a, b^\phi) - u(\varpi, b^\phi))u_{b^m}(\varpi, b^m)}{\phi^\phi_{b^m} - \frac{(u(a, b^\phi) - u(\varpi, b^\phi))u_{b^m}(\varpi, b^m)}{\delta (p - q)(v^m (\phi, b^m) - (u^\text{aut} + \delta \omega))}}.
\]

The denominator of Eq. (25) is negative since the right-hand side of Eq. (24) is increasing and concave in \( \phi \). Then, \( \phi^\phi_{b^m} \geq (<) 0 \) if and only if \( u_{b^m}(\varpi, b^m) \geq (<) 0 \). Hence, we conclude that it must be that \( u_{b^m}(\varpi, b^m) = 0 \) and \( \phi^\phi_{b^m} = 0 \), which implies that \( b^{m,e} = b^* \). Finally, the equilibrium level \( b^{m,e} \) is chosen so as to maximize Eq. (17) with respect to \( b^\phi \)
subject to $\phi = \phi^y (b^y, b^m)$ and $b^{m,e} = b^*$. Hence, $b^{y,e}$ solves:

$$u_{b^y} (\overline{\mu}, b^{y,e}) = -\delta (1 - q (1 - \phi^y (\cdot))) \left( \frac{q}{1-q(1-q^*(\cdot))} v^m (\phi^y (\cdot), b^{m,e}) - (u^{aut} + \delta \omega) \right) \phi_{b^y}^y. \quad (26)$$

By the implicit function theorem, Eq. (24) defines:

$$\phi_{b^y}^y = -\frac{\Delta (b^y)}{1 - \frac{u_{b^y} (\overline{\mu}, b^{y,e})}{u_{b^y} (\overline{\mu}, b^{m,e})} \phi_{b^y}^m}.$$

Then, $\phi_{b^y}^y < (>) 0$ if and only if $\Delta (b^y) > (<) 0$. Inspecting Eq. (26), we conclude that $u_{b^y} (\overline{\mu}, b^{y,e}) > (<) 0$ and, in turn, $b^{y,e} < (>) b^*$ when $\Delta (b^y) > (<) 0$. We now determine the condition under which constraint (18) is slack and constraint (19) is binding. This is true when $\delta < \delta^y$ where $\delta^y$ is the solution of the following equation:

$$\frac{u (a, b^{y,e} (\delta^y)) - u (\overline{\mu}, b^{y,e} (\delta^y))}{u (a, b^{m,e}) - u (\overline{\mu}, b^{m,e})} = \frac{v^m (\phi^y (b^y (\delta^y), b^{m,e}), b^{m,e}) - (u^{aut} + \delta \omega)}{\omega - \omega^*}.$$

Clearly, when $\delta = \delta^y$, $\phi = \phi^y (b^{y,e}, b^{m,e}) = \phi^m (b^{m,e})$ and both constraints (18) and (19) are simultaneously satisfied. □
References


