Social Macroeconomics
Working Paper Series

# Organisational ethics, narratives and social dysfunctions

SM-WP-2018
October 2018

**Dennis Snower**, Blavatnik School of Government, University of Oxford
**Steven Bosworth**, University of Reading

# Organizational Ethics, Narratives and Social Dysfunctions

Steven J. Bosworth and Dennis J. Snower

30 October 2018

**Abstract**

The paper explores the joint determination of economic output, wages, corporate culture, employees' ethical standards and monitoring intensity in an analysis of organizational dysfunction. The utility from economic activities can frequently be enhanced through unethical, socially harmful activity, such as corruption, sexual harassment and environmental degradation. The ethical sensitivities of managers and their employees are shaped through their social interactions and thus organizational dysfunctions can arise. Such dysfunctions may be mitigated through changes in government policies or social norms. These changes become particularly effective if they encourage the managers and employees to adopt more ethical narratives. This narrative shift gives the managers and employees more ethical objectives, guiding their economic behaviors. The more ethical objectives induce them to adopt even more ethical narratives, and so on, in a virtuous circle that promotes social welfare.

## 1 Introduction

This paper explores the joint determination of economic output, wages, corporate culture, employees' ethical standards and monitoring intensity in an analysis of organizational dysfunction. The underlying ideas are straightforward.

In markets that are neither perfectly regulated nor contractually complete, individuals' private utility from economic activities can frequently be enhanced through unethical, socially harmful activity, such as corruption, sexual harassment and environmental degradation. The resulting social dilemmas, whereby private returns exceed social returns, are mitigated through the ethical sensitivities of managers and their employees. These sensitivities, however, are shaped through the social interactions between managers and employees at the workplace.

If these sensitivities are deficient, organizations settle into a dysfunctional equilibrium, generating economic activity that does not maximize social welfare. Under these circumstances, managers are lax in their support of ethical behavior by their employees, and these employees consequently engage in socially harmful

activities with some impunity. The ethical laxity manifests itself in contractual incompleteness, as managers put little effort into checking that their employees are living up to the organization's alleged ethical standards. The employees take advantage of this ethical latitude. Such organizations tend to adopt lax ethical narratives, which assign correspondingly lax ethical identities to the managers and employees, encouraging them to value material payoffs over moral probity and to gain esteem from successfully cheating their customers and business partners.

This dysfunction may however mitigated through changes in government policies (such as increased supervision of ethical behavior, combined with punishments for malfeasance) and in social norms (supported by public commendation of ethical achievements and public outrage against unethical practices). These changes are likely to become particularly effective if they encourage the managers and employees to adopt more ethical narratives. Such narratives raise the managers' and employees' awareness of the social harm generated by the unethical activity. Thereby these narratives induce the managers to improve the corporate ethical culture, reduce the degree of contractual incompleteness, and use wage incentives to promote ethical behavior. In the same vein, the narratives induce employees to raise their ethical standards and accordingly reduce the level of unethical activity.

Furthermore, the shift toward more ethical narratives gives the managers and employees more ethical objectives, guiding their economic behaviors. The more ethical objectives induce them to adopt even more ethical narratives, and so on, in a virtuous circle that promotes social welfare. We derive conditions under which these beneficial feedback effects can materialize.

In short, this paper explores the following potentially important themes that have received little attention thus far in the economic literature on organizational behavior:

1. Organizational ethics arise through the interaction between managers and employees. Ethical standards affect the degree to which the production of economic output is promoted through unethical activity, and thereby influences the degree to which material prosperity is aligned with social welfare.

2. Both managers and employees are sensitive to ethical issues, alongside monetary payoffs. Their ethical sensitivities are generated by organizational narratives, which make sense of the roles that managers and employees play within the organization and create preference-shaping identities for managers and employees.

3. Managers' ethical behavior depends on their employees' responsiveness to ethical exhortations and to financial incentives. Managers do not attempt to reduce their employees' level of unethical activity if the cost of doing so, in terms of wages and lost output, exceeds their sensitivity to the associated social gain.

4. Employees' ethical behavior depends on their responsiveness to cognitive dissonance, the corporate ethical culture, the intensity of monitoring and the wage.

5. Thus managers' and employees' ethical behavioral are interdependent. In an "organizational equilibrium," these behaviors are consistent with one another. This equilibrium may be dysfunctional, in the sense that the organization's level of unethical behavior is socially suboptimal.

6. Managers and employees choose the narratives that enable them to reach their current objectives most effectively. However, since the narratives themselves influence these objectives, narrative choice is the outcome of a reflexive interaction between objectives and narratives.

7. Changes in government policies and in social norms can raise managers' and employees' sensitivities to unethical activity and thereby reduce organizational dysfunctions, aligning material prosperity more closely with social welfare.

Our analysis helps explain, first, the wide heterogeneity across organizations in their corporate ethical cultures and the degree of employee attachment to these cultures that is observed in practice (Alvesson, 2002; Cha and Edmonson, 2006). Second, our analysis explains why we observe high wages in organizations with strong corporate ethical culture to which employees are strongly attached (Mühlau and Lindenberg, 2003; Masakure and Gerhardt, 2016), while managers view wages and corporate culture enforcement as substitutes.[1] Third, our analysis helps account for the empirical regularity that government policies with an ethical thrust (such as many environmental policies) are effective primarily when reinforced through a complementary public ethical narrative. For example, recycling policies in the U.S. and Europe where relatively ineffective in the 1970s and 1980s, when a supporting ethical narrative was not widespread, but became far more effective since the 1990s, when narratives emphasizing the immorality of environmental degradation proliferated. Along similar lines, the U.S. prohibition policy in the interwar period had limited effectiveness and was eventually reversed since ethical narratives against alcohol consumption did not gain broad public support.

The rest of the paper is organized as follows. Section 2 covers essential underlying ideas. Section 3 presents our basic model of decision making, for a given narrative. Section 4 derives the associated organizational equilibrium. Section 5 focuses on the role government policies and narratives in promoting ethical behaviors. Finally Section 6 concludes.

[1] Green and Weisskopf (1990) show that the "worker disciplining" effect of efficiency wages varies greatly across industries. Industries characterized as "secondary," with high turnover and low identity-building tend to rely most on wages to incentivize workers.

# 2 Underlying Ideas

Selfish optimizers with unique, internally consistent objectives – such as those in conventional neoclassical economic analysis – will invariably engage in unethical activities, provided that the expected net private return is positive. In this context, the job of ensuring ethical behavior falls to the government, through its regulations and incentives (rewards and punishments, such as subsidies for ethical practices and fines for malfeasance). In our analysis, by contrast, managers and employees have ethical sensitivities alongside their material needs and wants, and these sensitivities arise from the interaction of managers and employees in the production process. Thus preferences in our analysis are not unique to individual agents, but instead are the outcome of the interplay between individuals and their social context. This preference-generating interplay is shaped by the narratives adopted by the managers and employees, defining their identities within the organization.

## 2.1 Motives and Ethical Sensitivities

The psychological mechanism whereby preferences arise from social interactions, shaped by identity-creating narratives, is a well-known insight from motivation psychology,[2] namely, that all human behavior is motivated and that people have access to multiple, discrete motives. Each of these motives can be associated with a distinct objective function.[3] "Motives" – in the sense that the term is used in motivation psychology – are forces that give direction and energy to one's behavior, thereby determining the objective of the behavior, as well as its intensity and persistence.[4] Motives that are associated with heightened ethical sensitivities are Care (seeking to promote the well-being of others) and Affiliation (seeking belonging within social groups).[5] By contrast, motives that are linked to low ethical sensitivities are Status-Seeking (seeking social standing and social influence),[6] Anger (aggressive responses to threats)[7] and Fear (defensive responses to threats).[8] Each of these motives is associated with a different set of preferences. In our analysis, we focus exclusively on the role of motives in shaping ethical sensitivities.

Which motive is active in an individual at any particular time depends on the individual's environment. Our analysis focuses exclusively on the individual's social environment, as defined by a narrative. A given narrative interprets

[2] See, for example, the survey by Heckhausen and Heckhausen (2008).

[3] See, for example, Bosworth, Snower and Singer (2016) and Snower and Bosworth (2016).

[4] See Elliot and Covington, 2001; following Atkinson, 1964.

[5] Care is concerned with nurturance, compassion, and care-giving, e.g. Weinberger et al., (2010). This motive is often distinguished from the Affiliation motive, e.g. McDougall (1932), Murray's (1938), McAdams (1980), Heckhausen and Heckhausen (2008).

[6] For example, Heckhausen and Heckhausen (2008).

[7] See, for example, McDougall's (1932) concept of anger/rage, Murray's (1938) aggression and defendance, Heckhausen's (1989) aggression, and Reiss' (2004) vengeance.

[8] See, for example, the concept of threat avoidance in McDougall (1932) and Murray (1938). Note, however, that Fear may lead not only to non-cooperation, but under some circumstances also to cooperation (e.g. Taylor (2006)).

the individual's social environment in a particular way, thereby shaping the individual's social identity. This leads to the activation of particular motives, which in turn generate particular ethical sensitivities.

Given these ethical sensitivities, the managers in our analysis make decisions concerning the corporate culture, the level of monitoring to engage in, and the wage; whereas the employees in our analysis make decisions concerning their identification with ethical standards and the level of unethical activity. In making their decisions, managers face a tradeoff between their awareness of the social harm from the unethical activity and the cost of monitoring this activity. Employees face a tradeoff between the private return from the unethical activity and the cognitive dissonance generated by the disparity between their avowed ethical standards and their actions.

Specifically, employees face two types of dissonance:

- *principle-action dissonance:* a discrepancy between their internal moral principles and their actions in the social dilemma activities and
- *manager-employee principle dissonance:* a discrepancy between the manager's overt moral principles and employees' internal moral principles.

Employees seek to avoid these two types of dissonance at their workplace. The manager can affect the magnitude of employees' dissonance through the corporate culture, monitoring and wages. The corporate culture is described by the organization's ethical standard and the degree to which the manager promotes this standard. The employee sets her personal ethical standard between the organization's standard and her unethical activities, with the aim of minimizing dissonance. The more the organization's standard diverges from her personal standard and the more often the manager promotes the organization's standard, the greater is her manager-employee principle dissonance. The more her personal standard diverges from her actions regarding her unethical activity, the greater is her principle-action dissonance. If these two types of dissonance become sufficiently large, the employees no longer attempt to comply with the corporate culture, thereby risking contract termination.

In the organizational equilibrium, the manger's and employee's decisions are consistent with one another. This equilibrium may be dysfunctional.

## 2.2 Narratives and Identities

Narratives[9] are sequences of causally linked events, particularly ones linking people's actions to consequences, which may be used as a template for understanding our ongoing experiences.[10] Narratives provide simple mental models that help us understand our social environment. They focus our attention on particular events and causal relations, whereby we predict further events. Importantly, narratives activate distinctive motives in us. They do so by assigning

[9] Akerlof and Snower (2016) describe these and other functions of narratives.
[10] See, for example, Bruner, 1991.

social roles to people, placing them into well-defined relationships to one another. These social roles are often endowed with normative force. Thereby narratives shape our social identities, which in turn influence our objectives within organizations, along lines investigated profoundly by Akerlof and Kranton (2005). As noted, this paper focuses on the influence of narratives and identities on people's ethical sensitivities.

New organizational narratives interpret mangers' and employees' social setting in new ways, giving these agents new social roles within a broader story, and thereby bestowing them with new identities. These new identities motivate them to behave in new ways. In particular, new ethical narratives generate new ethical sensitivities. Our analysis indicates that narrative shifts can be set in motion through changes in government policies and changes in social norms. Thereby narrative shifts can induce changes in the organizational equilibrium.

# 3 The Decision Making Problems

We assume that a manager employs a fixed number of identical employees indexed by $i$. Each employee provides one unit of standard labor input to generate one unit of production. In addition, the employee can engage in an unethical activity $\tau \in [0, \tau^{max}]$. We may think of the unethical activity as a misconduct such as sexual harassment: it generates no marketable output, a positive private utility for the miscreant employee and a negative social payoff which the manager (partially) internalizes. Alternatively, it could involve financial fraud or falsification of research results, which may contribute to the marketable output of the organization (as long as they remain undetected) while generating a positive private payoff to the employee (bonuses, promotions) and a negative social payoff.

We conceive of the interaction between managers and employees as a two-period game. In the first period, the manager sets an ideal ethical standard, a threshold level of behavior above which employees are fired, the wage, and monitoring intensity. In the second stage, employees decide how much to internalize the organizational culture, whether to comply with the requested level of the activity. The organizational culture, level of the activity, and resulting payoffs in terms of wages constitute a subgame-perfect Nash equilibrium of this game.

Let there be $j$ narratives: $N_j$, with higher $j$ representing higher social awareness of the social harm from the unethical activity. The unethical activity provides a positive private (psychic or material) payoff to the employee ($\beta_j \tau - \frac{\rho}{2}\tau^2 > 0$, where $\beta_j,\ \rho > 0$) and a negative social payoff ($-\Lambda\tau$, where $\Lambda > 0$ is a constant). More ethically demanding narratives (with higher $j$) make the employee more aware of the social harm from his unethical activity and thus are associated with lower $\beta_j$.

The manager's social awareness is captured by her disutility from employee $i$'s social dilemma activity: $\sigma_j \tau$, where $\Lambda \geq \sigma_j \geq 0$, so that when $\sigma_j = 0$, the manager has no awareness, whereas when $\sigma_j = \Lambda$, the manager has full awareness.

Each employee receives the wage $w \geq 0$. The manager observes the socially undesirable activity $\tau_i$ with probability $p$. This probability is determined by the cost of monitoring $c_m$. The manager is able to write contracts under which $i$ may be fired if $\tau_i > \tau^*$ for some contractually specified $\tau^*$. A higher $p$ entails more contractual completeness, but comes at a greater cost $c_m$. Thus $p$ will be called the "monitoring intensity."

The manager's objective function under narrative $j$ is

$$V_j = 1 - w - \sigma_j \tau_i - \frac{c_m}{2} p^2 \tag{1}$$

where the payoff from the marketable output is unity.

employees experience dissonance arising from a disparity between their internal moral principles and their actions, as well as a disparity between their internal moral principles and the manager's principles. In the spirit of Rabin (1994), we assume that agents adopt moral standards flexibly to reduce the cognitive dissonance associated with failing to act in accordance with what they consider to be right. In particular, each employee $i$ maintains an internal moral standard $\mu_i$ which cannot be observed, and is asked to adhere to the manager's moral standard, $\nu^* \geq 0$, chosen costlessly be the manager. In this context, employees experience two types of dissonance:

- Principle-action dissonance arises from a discrepancy between the employee's principle (the internal moral standard $\mu_i$) and the employee's action (the socially harmful activity $\tau$).
- Inter-principle dissonance arises from a discrepancy between the manager's principle (the organization's overt moral standard $\nu^*$) and the employee's principle (the internal moral standard $\mu_i$).

The manager chooses the frequency $q \in [0, 1]$ with which the employees are reminded of the organization's moral standard $\nu^*$. The variable $q$ is called the organization's "corporate culture." If employees reject the corporate culture, they free themselves of inter-principle dissonance, but they also run the risk of being fired.

The employee faces a straightforward tradeoff. She seeks the private return from the social dilemma activity $\tau$, but this activity generates dissonance that the employee wishes to avoid. As shown below, the organization's moral standard $\nu^*$ will be so strict that it is irreconcilable with any engagement in the social dilemma activity high. If the employee chooses to engage in this activity nonetheless ($\tau_i > 0$), then she will experience dissonance. The higher is the employee's internal moral standard ($\mu_i$), the larger will be the principle-action dissonance ($\mu_i - \tau_i$) and the smaller will be the inter-principle dissonance ($\nu^* - \mu_i$).

Thus the employee $i$'s utility under narrative $j$ may be expressed as

$$U_{ij} = \begin{cases} w + \beta_j \tau_i - \frac{\rho}{2}\tau_i^2 - \frac{\alpha}{2}\left(\max\left\{\tau_i - \mu_i,\, 0\right\}\right)^2 - q\gamma\frac{\alpha}{2}\left(\max\left\{\mu_i - \nu^*,\, 0\right\}\right)^2 & \tau_i \leq \tau^* \\ (1-p)\left(w + \beta_j \tau_i - \frac{\rho}{2}\tau_i^2 - \frac{\alpha}{2}\left(\max\left\{\tau_i - \mu_i,\, 0\right\}\right)^2 - q\gamma\frac{\alpha}{2}\left(\max\left\{\mu_i - \nu^*,\, 0\right\}\right)^2\right) & \tau_i > \tau^* \\ +p\left(\beta_j \tau_i - \frac{\rho}{2}\tau_i^2 - \frac{\alpha}{2}\left(\max\left\{\tau_i - \mu_i,\, 0\right\}\right)^2\right) & \end{cases}. \tag{2}$$

where $\alpha$ is a positive constant, denoting the employee's sensitivity to both types of dissonance, and $\gamma$ represents the size of the inter-principle dissonance relative to the principle-action dissonance.

Note that if the employee engages in less than the threshold level $\tau^*$ of the social dilemma activity, she earns the wage $w$ and the private return $\tau_i$, while paying the psychic cost for the principle-action dissonance ($\frac{\alpha}{2}\left(\max\left\{\tau_i - \mu_i,\, 0\right\}\right)^2$) and the employer-employee dissonance ($q\gamma\frac{\alpha}{2}\left(\max\left\{\mu_i - \nu^*,\, 0\right\}\right)^2$) . On the other hand, if she engages in more than the threshold level $\tau^*$ of activity $\tau$, she retains her position only with probability $(1-p)$. With probability $p$, she gets caught and has to forfeit her wage $w$ (as in a standard imperfect contracting model). There are also psychic benefits to leaving the organization, since she would be free from the employer-employee dissonance.

For the objective functions above, the manager's control variables are the wage $w$, the intensity of monitoring $p$, and the corporate culture $q$. The employee's control variables are the level of the socially harmful activity $\tau$ and the employee's moral standard $\mu$.

# 4 The Organizational Equilibrium

In the organizational equilibrium, the manager's optimal choices are consistent with the employee's optimal choices. To identify this equilibrium, we first find the contractually binding level $\tau^*$ of the social dilemma activity and the level of the employee's internal moral standard $\mu_i^*$ which are incentive-compatible. Next, given these incentive compatible levels, we derive the manger's decisions concerning the equilibrium wage $w^*$, the corporate culture $\nu^*$ and $q^*$, and the intensity of monitoring $p^*$.

## 4.1 Incentive Compatibility Conditions

At this incentive-compatible level of $\tau^*$, the employee is indifferent between choosing $\tau^*$ and her optimal choice of $\tau_i = \widehat{\tau}$ at which she does not comply with $\tau^*$:

$$\widehat{\tau} = \arg\max_{\tau_i} (1-p)\left(w + \beta_j \tau_i - \frac{\rho}{2}\tau_i^2 - \frac{1}{2}\left(\tau_i - \mu_i\right)^2 - q\cdot\frac{1}{2}\left(\mu_i - \nu^*\right)^2\right) + p\left(\beta_j \tau_i - \frac{\rho}{2}\tau_i^2 - \frac{1}{2}\left(\tau_i - \mu_i\right)^2\right).$$

By the first-order condition,[11]

$$\widehat{\tau} = \frac{\alpha\mu_i + \beta_j}{\alpha + \rho}. \tag{3}$$

Thus the incentive compatibility condition, at which the employee is indifferent between compliance and non-compliance, is

$$\begin{aligned} & w + \beta_j\tau^* - \frac{\rho}{2}\tau^{*2} - \frac{1}{2}\left(\tau^* - \mu_i\right)^2 - q\cdot\frac{\gamma}{2}\left(\mu_i - \nu^*\right)^2 \\ &= (1-p)\left(w + \beta_j\widehat{\tau} - \frac{\rho}{2}\widehat{\tau}^2 - \frac{1}{2}\left(\widehat{\tau} - \mu_i\right)^2 - q\cdot\frac{1}{2}\left(\mu_i - \nu^*\right)^2\right) + p\left(\beta_j\widehat{\tau} - \frac{\rho}{2}\widehat{\tau}^2 - \frac{1}{2}\left(\widehat{\tau} - \mu_i\right)^2\right) \end{aligned}$$

which implies the following value of the threshold level $\tau^*$:

$$\tau^* = \frac{\alpha\mu_i + \beta_j - \sqrt{p\left(\alpha+\rho\right)\left(2w - \alpha q\gamma\left(\mu_i - \nu^*\right)^2\right)}}{\alpha + \rho}. \tag{4}$$

Agent $i$ also chooses her internal moral standard $\mu_i = \mu^*$ which minimizes her dissonance under $\tau^*$. Partially differentiating $U_{ij}$ with respect to $\mu_i$, we derive the following first-order condition

$$\alpha\left(\tau^* + q\gamma\nu^* - \mu^*\left(q\gamma + 1\right)\right) = 0$$

which implies that that

$$\mu^* = \frac{q\gamma\nu^* + \tau^*}{q\gamma + 1} \tag{5}$$

## 4.2 The Manager's Decision Problem

We now consider the manager's problem from the standpoint of setting $w$, $\nu$, $q$ and $p$ optimally. Recall that the manager's problem is to maximize the objective function (1): $V_j = 1 - w - \sigma_j\tau_i - \frac{c_m}{2}p^2$.

Substituting $\tau_i = \tau^*$ into the managerial objective function (1) and differentiating $V_j$ with respect to $w$, we obtain the following expression for the wage:[12]

$$w^* = \frac{\alpha q\gamma\left(\beta_j - \nu\rho\right)^2 + p\sigma_j^2\left(\alpha+\rho\right)\left(q\gamma+1\right)^2}{2\left(\alpha^2 q\gamma\left(q\gamma + p\right) + \alpha q\gamma\rho\left(2\left(q\gamma+1\right)+p\right) + \rho^2\left(q\gamma+1\right)^2\right)}. \tag{6}$$

In order to derive the equilibrium organizational culture $\nu^*$ and $q^*$, we take the partial derivative of $V_j$ with respect to $\nu$ and then substitute in the equilib-

[11] $\alpha\mu_i + \beta_{1j} - \widehat{\tau}\left(\alpha + \beta_j\rho\right) = 0$.

[12] The first-order condition of the manager's problem with respect to the wage is $\frac{p\sigma_j(\alpha+\rho)(q\gamma+1)}{\sqrt{p(\alpha+\rho)\left(2w\left(\alpha^2 q\gamma(q\gamma+p)+\alpha q\gamma\rho(2(q\gamma+1)+p)+\rho^2(q\gamma+1)^2\right)-\alpha q\gamma\left(\beta_j-\nu\rho\right)^2\right)}} - 1 = 0$, from which the equation equation is derived.

rium wage $w = w^*$ to derive the equilibrium organizational norm:[13]

$$\nu^* = \frac{\rho\left(\beta_j - \sigma_j\right) - \sigma_j\left(\alpha + \rho\right)\left(q\gamma + p\right)}{\rho^2}. \tag{7}$$

Likewise taking the derivative of $V_j$ with respect to $q$ and then substituting $w^*$, $\nu^*$ we see that

$$\frac{\partial V_j}{\partial q} = \frac{\alpha\gamma\sigma_j^2}{2\rho^2} > 0, \tag{8}$$

meaning that the manager sets the frequency of moral reminders $q^* = 1$ as high as possible regardless of where the organizational norm lies.

Finally, maximizing the manager's objective with respect to monitoring probability (and again plugging $w^*$, $\nu^*$ and $q^*$ into the resulting first-order condition[14]), we obtain the equilibrium monitoring intensity:

$$p^* = \frac{\sigma_j^2\left(\alpha + \rho\right)}{2\rho^2 c_m}, \tag{9}$$

The organizational equilibrium is thus described by the equilibrium social dilemma activity equation (**??**), the equilibrium frequency of moral reminders ($q^* = 1$), the wage equation (6), the equilibrium corporate culture equation (7), and the equilibrium monitoring intensity equation (9). Plugging $p^*$ back into the expressions for $w^*$ and $\nu^*$ we see that

$$w^* = \frac{\left(\alpha + \rho\right)^2\sigma_j^4 + 2\alpha c_m\gamma\sigma_j^2\rho^2}{4\rho^4 c_m}, \tag{10}$$

$$\nu^* = \frac{\beta_j - \sigma_j\left(1 + \gamma\right)}{\rho} - \frac{\sigma_j\alpha\gamma}{\rho^2} - \frac{\sigma_j^3\left(\alpha + \rho\right)^2}{2\rho^4 c_m} \tag{11}$$

and

$$\tau^* = \frac{\beta_j}{\rho} - \frac{\sigma_j^3\left(\alpha + \rho\right)^2}{2c_m\rho^4} - \frac{\alpha\gamma\sigma_j}{\rho^2}. \tag{12}$$

Organizational cultures less permissive than $\nu^*$ will backfire in the sense that, for given wages, employees will revert to the behavior $\hat{\tau}$ since the incentive-compatibility condition (Eq. 4) is violated. Here the manager could consider to increase $w$, in order to induce greater compliance with the organization's moral standard. However, this course would be sub-optimal, since the manager's willingness to pay for more ethical behavior would be less than what the employees need to perform more ethical behavior. In other words, more exacting organizational cultures are infeasible since employees are "not being paid enough" to

[13] The first-order condition of the manager's problem with respect to $\nu$ is $\frac{\alpha q\gamma\left(\rho\left(\beta_j - \nu\rho\right) - \sigma_j(\alpha+\rho)(q\gamma+p)+\rho\right)}{\alpha^2 q\gamma(q\gamma+p)+\alpha q\gamma\rho(2(q\gamma+1)+p)+\rho^2(q\gamma r+1)^2} = 0$, from which the equilibrium corporate culture is derived.

[14] The first-order condition of the manager's problem with respect to $p$ is $\frac{\sigma_j^2(\alpha+\rho)}{2\rho^2} - c_m p = 0$, from which the equilibrium monitoring intensity is derived.

| | $\tau^*$ | $w^*$ | $\nu^*$ | $p^*$ |
|---|---|---|---|---|
| $\sigma_j$ | $-$ | $+$ | $-$ | $+$ |
| $-\beta_j$ | $-$ | $-$ | $-$ | $-$ |

Table 1: Effects of Government Policies

identify them. Since the employees consider their outside option as entailing lower tension between their behavior and the manager's standards, the function of the wage is partially to get them to buy in to the organizational culture $\{\nu^*, q^*\}$.

Furthermore, the equilibrium above implies that organizational culture and wages are inversely related in the organizational equilibrium:

$$\frac{dw^*}{d\nu} = -\frac{\alpha\gamma\rho\left(\beta_j - \rho\nu\right)}{\alpha^2\gamma(q\gamma + p) + \alpha\gamma\rho\left(2\left(\gamma + 1\right) + p\right) + \rho^2\left(\gamma + 1\right)^2} < 0 \text{ at } \nu^* \tag{13}$$

implying that monetary incentives and organizational culture are substitutes for the manager. From the employee's perspective however, organizational culture and incentives are complementary. Recall that the employee's internal moral (how much theft, bribe-taking, etc.) standard depends negatively on the wage:

$$\frac{d\mu^*}{dw} = -\frac{1}{\sigma_j\left(1 + \gamma\right)} < 0 \text{ at } w^*. \tag{14}$$

This means that while managers can use organizational culture to economize on wages, ultimately employees identify more with organizational cultures when they are paid more.

# 5 Government Policies

In this context, consider two government policies: (1) a fine that increases the cost to the manager from the unethical activity $\tau$ and (2) a fine that increases the employee's cost of this activity. The manager-oriented policy raises the manager's sensitivity $\sigma_j$, whereas the employee-oriented policy reduces the employee's payoff $\beta_j$. The qualitative effects are summarized in Table 1. Note that these two policies have quite different effects on the organization's activities.

- The manager-oriented policy leads to an improvement in the corporate ethical culture ($\nu^*$), while also increasing the wage and the intensity of monitoring. Consequently this policy leads to a reduction in the unethical activity ($\tau^*$).

- The employee-oriented policy also leads the manager to improve the corporate ethical culture ($\nu^*$), since higher ethical standards are now easier to achieve. In addition, however, the manager reduces the monitoring intensity ($p^*$), since the employees' greater sensitivity to the social harm

from the unethical activity makes it less worthwhile to observe the unethical activity with such frequency. The manager also reduces the wage ($w^*$), since the employee no longer needs such a high wage to induce more ethical behavior. These changes also lead to a reduction in the unethical activity ($\tau^*$), but the effect is weaker (per unit of the fine) than for the manager-oriented policy.

In short, the manager-oriented policy is a more effective way of mitigating organizational dysfunction than the employee-oriented policy, even though the application of conventional principles of justice would suggest that the employee – the perpetrator of the misdeed – should receive the fine.

# 6 Narrative Shifts

Under what conditions can these policies induce the manager to choose a more ethical narrative, i.e. one that makes both the manager and the employees more sensitive to the social harm from unethical activity?

The manager faces a tradeoff. On the one hand, if the employees accept the new narrative, then she would benefit, since the employees would engage in less unethical behavior on account of their increased sensitivity (lower $\beta_j$). On the other, such a narrative shift would also entail that the manager herself becomes more sensitive to the social harm from whatever unethical activity is still being performed (higher $\sigma_j$). In deciding whether to choose a more ethical narrative, the manager needs to consider whether the resulting ethical sensitivity of the employees rises sufficiently to induce the manager to take ethical breaches more seriously.

Let us decompose the manager's ethical sensitivity in the following way:

$$\sigma_j = \sigma_0 + \lambda_j^m + \pi^m - \eta\left(\lambda_j^e + \pi^e\right) \tag{15}$$

where $\lambda_j^m + \pi^m + \eta\left(\lambda_j^e + \pi^e\right)$ is the manager's self-recognized level of the social harm from activity $\tau$, $\pi^m$ represents exogenous regulatory costs incident on the manager (e.g. fines or reputational losses from corrupt behavior), $\pi^e$ represents regulatory costs incident on the employee, and $\sigma_0$ are the manager's pecuniary private costs from the activity. The parameter $\eta \geq 0$ is the "narrative multiplier". This links the acknowledged social harms that the employees and manager recognize from the activity since they must both use the same language to talk about it. When $\eta < 1$, this represents the ethical sensitivities of one party acting as a substitute for those of the other party, whereas for $\eta > 1$, the sensitivities of the two parties are complementary, as they have a strong reason to align their narratives.

Similarly, we can decompose the employee's (net) marginal utility benefit from $\tau$, $\beta_j$, as follows

$$\beta_j = \beta_0 + \lambda_j^e + \pi^e + \eta\left(\lambda_j^m + \pi^m\right)$$

with the term $\beta_0 > 0$ representing the employee's private utility from activity $\tau$ in the absence of any social or economic incentives and $\lambda_j^e + \pi^e + \eta\left(\lambda_j^m + \pi^m\right)$ representing the employee's acknowledged social harm. Note that there is some element of "double counting" inherent in narratives: if the manager faces private regulatory costs from the activity then she might as well speak of it as more socially harmful (since she now has to care about it this much more). This has the added benefit of increasing the employee's acknowledged level of social harm. The terms $\lambda_j^m$ and $\lambda_j^e$ represent the flexible components of the manager's and employee's narratives, respectively.

The interdependence in the recognized social harm from the activity between the manager and the employee creates a "narrative equilibrium". In the longer run, we suppose that this narrative equilibrium evolves in such a way as to satisfy the first-order conditions

$$\frac{dV_j}{d\lambda_j^m} = \frac{dU_j}{d\lambda_j^e} = 0.$$

Or, in other words, people adopt the narratives which suit them best. Solving first for the manager's first-order condition we obtain

$$\lambda_*^m = \frac{\beta_0\rho - \left(\alpha\eta\gamma + \rho\left(1+\eta^2\right)\right)\left(\lambda_j^e + \pi^e\right)}{\alpha\gamma + 2\eta\rho} - \sigma_0 - \pi^m \tag{16}$$

which we will call the *manager narrative compatibility condition* (MC).

Policies affecting the employee's costs of engaging in activity $\tau$ can shift the manager's narrative, and the direction of this relationship depends on the size of the narrative multiplier:

$$\frac{d\sigma_*}{d\pi^e} = \frac{\left(\eta^2 - 1\right)\rho}{\alpha\gamma + 2\eta\rho}.$$

We can use $\lambda_*^m$ to evaluate the employee's first-order condition and derive the equilibrium narrative:

$$\lambda_*^e = \frac{3\left(\eta\sigma_0 - \beta_0\right) + 3\left(\eta^2 - 1\right)\pi^e + \sqrt{6}\sqrt{\frac{c_m(\alpha\gamma+2\eta\rho)^2\left(\sqrt{\eta(\alpha^2\eta\gamma^2 - \alpha\gamma\rho(3\eta\gamma+4\eta+3) - 2\eta\rho^2)} - 2\alpha\eta\gamma + \eta\rho\right)}{\eta(\alpha+\rho)^2}}}{3\left(\eta^2 - 1\right)}.$$

we call this employee narrative compatibility condition, EC.

The resulting equilibrium is pictured in Figure 1. In the long-run narrative equilibrium, narratives underlying the ethical sensitivities of the manager and employee are consistent with one another. Such an equilibrium is pictured by Point A in the figure.

The manager-oriented policy above may be pictured by a rightward shift in MC. The immediate impact of the policy is to raise the manager's ethical sensitivity ($\sigma_j$) for any given level of the employee's social awareness $\beta_j$. But the rise in the manager's ethical sensitivity ($\sigma_j$) may lead the employee to choose a more ethical narrative, leading to a fall in $\beta_j$. This in turn leads the manager to choose a more ethical narrative, leading to a further rise in $\sigma_j$, and so on, until the new narrative equilibrium at Point B is reached.

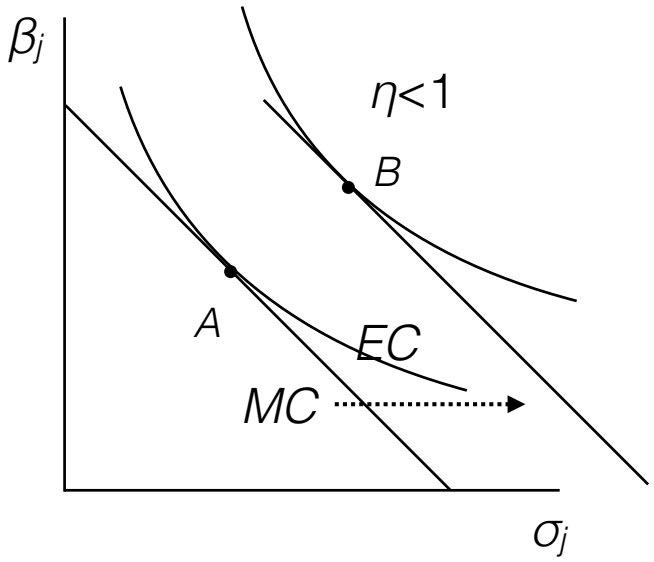


Figure 1: Narrative equilibrium with $\eta < 1$

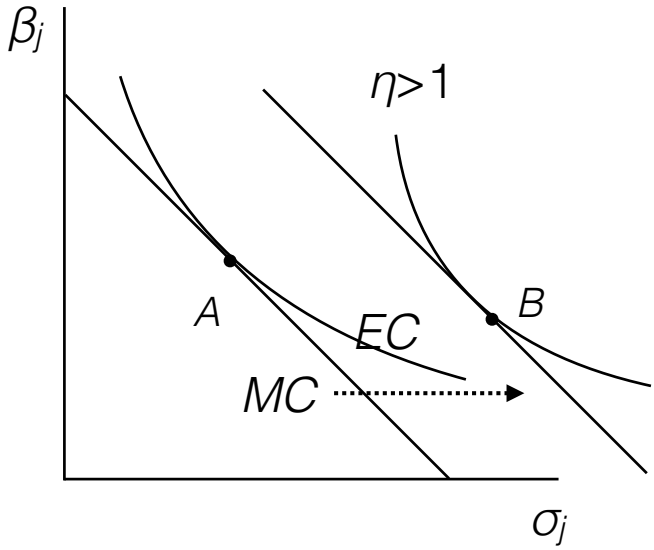


Figure 2: Narrative equilibrium with $\eta > 1$

# 7 Concluding Remarks

Transforming organizational cultures is a matter of crucial importance in today's society. Many workplaces for example are grappling with cultures of sexual harassment and taking long-overdue steps to root out instances of abuse. We show that this has both direct effects and an indirect reinforcing effect through organizational cultural change. Another key application of our results is to combatting research misconduct within academia. Given the high monitoring costs and strong motivation to deter bad practice in this domain, we explain very strong cultural norms against malpractice coinciding with high wages for established researchers. One implication here is that performance incentives, such as those under the U.K.'s Research Excellence Framework, may actively harm cultural protections against malpractice.

Our results also highlight the limits and tradeoffs in effecting organizational change. Organizational culture is constrained by the manager needing "buy-in" from employees for any changes. This could be one reason why job satisfaction is correlated with strong ethics across organizations (Koh and Boo, 2001): Employees who are attached to their organization can be held to high standards.

# References


[1] Akerlof, G. A. & Kranton, R. E. (2005). Identity and the economics of organizations. Journal of Economic Perspectives, 19(1), 9-32.

[2] Akerlof, G. A. & Snower, D. J. (2016). Bread and bullets, Journal of Economic Behavior and Organization, 126(B), 72-94.

[3] Alvesson, M. (2002). Understanding Organizational Culture. London: SAGE Publications.

[4] Atkinson, J. W. (1964). An Introduction to Motivation. Princeton: Van Nostrand.

[5] Bosworth, S. J., Singer, T., & Snower, D. J. (2016). Cooperation, motivation and social balance. Journal of Economic Behavior and Organization, 126, 72-94.

[6] Bruner, J., 1991. The narrative construction of reality. Critical Inquiry, 18, 1-21.

[7] Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. Educational Psychology Review, 13(2), 73-92.

[8] Green, F. & Weisskopf, T. E. (1990). The worker discipline effect: A disaggregative analysis. Review of Economics and Statistics, 72(2), 241-249.

[9] Heckhausen, J., and H. Heckhausen (2008). Motivation and Action. Cambridge: Cambridge University Press.

[10] Koh, H. C. & Boo, E. H. Y. (2001). The link between organizational ethics and job satisfaction: A study of managers in Singapore. Journal of Business Ethics, 29(4), 309-324.

[11] Masakure, O. & Gerhardt, K. (2016). Employee commitment and wages in the private sector. Labour, 30(1), 38-60.

[12] Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. Journal of Marketing Research, 45(6), 633-644.

[13] McAdams, D. P. (1980). A thematic coding system for the intimacy motive. Journal of Research in Personality, 14(4), 413-432.

[14] McDougall, W. (1932). The Energies of Men. London: Methuen.

[15] Mühlau, P. & Lindenberg, S. (2003). Efficiency wages: Signals or incentives? An empirical study of the relationship between wage and commitment. Journal of Management and Governance, 7(4), 385-400.

[16] Murray, H. A. (1938). Explorations in Personality. New York: Oxford University Press.

[17] Rabin, M. (1994). Cognitive dissonance and social change. Journal of Economic Behavior & Organization, 23(2), 177-194.

[18] Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. Review of General Psychology, 8(3), 179-193.

[19] Schwarz, S. H. (2012). An overview of the Schwartz theory of basic values. Online Readings in Psychology and Culture, 2(1).

[20] Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proceedings of the National Academy of Sciences of the United States of America, 109(38), 15197-15200.

[21] Snower, D. J. & Bosworth, S. J. (2016). Identity-driven cooperation versus competition. American Economic Review: Papers and Proceedings, 106(5), 420-424.

[22] Taylor, S. E. (2006), Tend and befriend: Biobehavioral bases of affiliation under stress. Current Directions in Psychological Science, 15(6), 273-277.

# Appendix 1 – Performance vs. ethics

Many times agents commit corrupt or dishonest behavior because they are pursuing other objectives which the manager has incentivized. While it may be reasonable to assume that managers prefer less corruption *ceteris paribus*, in

equilibrium they may recognize a tradeoff with other performance objectives. We capture this with the following extensions to the model. Firstly, we suppose that in addition to $\tau$, managers care about employees meeting a deliverable performance target $d$:

$$V_j = 1 - w - \sigma_j \tau_i - \frac{c_m}{2} p^2 + \phi \log(d_i)$$

Here the parameter $\phi$ captures how much the manager cares about the deliverable performance objective $d_i$. Agents have a disutility of performing at high levels, and engaging in (manager undesirable) activity $\tau$ may reduce this disutility (concretely, agents may meet performance targets more easily by engaging in fraud). We capture this in the agent's utility by the cost by $-\delta \cdot d_i$, with $\delta = \delta_0 - \delta_\tau \cdot \tau_i$. That is, the higher $\tau$ is, the less disutility from meeting performance objective $d_i$.

We assume that agent $i$'s performance is easily observed and contracted upon; so the manager can ask for a level of performance $d^*$ which leaves $i$ indifferent between complying and quitting.

Agent $i$'s utility therefore becomes

$$U_{ij} = \begin{cases} w + \beta_j \left(\tau_i - \frac{\rho}{2}\tau_i^2\right) - \frac{\alpha}{2}\left(\max\{\tau_i - \mu_i, 0\}\right)^2 - q\gamma\frac{\alpha}{2}\left(\max\{\mu_i - \nu^*, 0\}\right)^2 - \delta d^* & \tau_i \leq \tau^* \\ (1-p)\left(w + \beta_j \left(\tau_i - \frac{\rho}{2}\tau_i^2\right) - \frac{\alpha}{2}\left(\max\{\tau_i - \mu_i, 0\}\right)^2 - q\gamma\frac{\alpha}{2}\left(\max\{\mu_i - \nu^*, 0\}\right)^2 - \delta d^*\right) & \\ \quad + p\left(\beta_j \left(\tau_i - \frac{\rho}{2}\tau_i^2\right) - \frac{\alpha}{2}\left(\max\{\tau_i - \mu_i, 0\}\right)^2 - \delta d^*\right) & \tau_i > \tau^* \end{cases}. \tag{17}$$

## A1.1 The indifference condition

We start as before by finding the contractually binding level of $\tau^*$ which is incentive-compatible. This means that agents are indifferent between choosing $\tau^*$ and their optimal choice of $\tau_i = \widehat{\tau}$ were they to choose not to comply with $\tau^*$. We find $\widehat{\tau}$ by

$$\begin{aligned}\widehat{\tau} = \arg\max_{\tau_i} (1-p)\left(w + \beta_j\left(\tau_i - \frac{\rho}{2}\tau_i^2\right) - \frac{\alpha}{2}\left(\tau_i - \mu_i\right)^2 - q\cdot\frac{\gamma}{2}\left(\mu_i - \nu^*\right)^2 - \delta d^*\right) \\ + p\left(\beta_j\left(\tau_i - \frac{\rho}{2}\tau_i^2\right) - \frac{\alpha}{2}\left(\tau_i - \mu_i\right)^2 - \delta d^*\right).\end{aligned}$$

The first-order condition is $\alpha\mu_i + \beta_j + \delta_\tau d^* - \widehat{\tau}(\alpha + \beta_j\rho) = 0$, giving us

$$\widehat{\tau} = \frac{\alpha\mu_i + \beta_j + \delta_\tau d^*}{\alpha + \beta_j \rho}.$$

We can now solve the incentive compatibility condition to find $\tau^*$:

$$\begin{aligned}&w - \beta_j\left(\tau^* - \frac{\rho}{2}\tau^{*2}\right) - \frac{\alpha}{2}\left(\tau^* - \mu_i\right)^2 - q\cdot\frac{\gamma}{2}\left(\mu_i - \nu^*\right)^2 - \left(\delta_0 - \delta_\tau\cdot\tau^*\right)d^* \\ &= (1-p)\left(w + \beta_j\left(\widehat{\tau} - \frac{\rho}{2}\widehat{\tau}^2\right) - \frac{\alpha}{2}\left(\widehat{\tau} - \mu_i\right)^2 - q\cdot\frac{\gamma}{2}\left(\mu_i - \nu^*\right)^2 - \left(\delta_0 - \delta_\tau\cdot\widehat{\tau}\right)d^*\right) \\ &+ p\left(\beta_j\left(\widehat{\tau} - \frac{\rho}{2}\widehat{\tau}^2\right) - \frac{\alpha}{2}\left(\widehat{\tau} - \mu_i\right)^2 - \left(\delta_0 - \delta_\tau\cdot\widehat{\tau}\right)d^*\right)\end{aligned}$$

giving us

$$\tau^* = \frac{\alpha\mu_i + \beta_j + \delta_\tau d^* - \sqrt{p\left(\alpha + \beta_j\rho\right)\left(2w - \alpha q\gamma\left(\mu_i - \nu^*\right)^2\right)}}{\alpha + \beta_j\rho}.$$

Agent $i$ also chooses $\mu_i = \mu^*$ which most effectively reduces her cognitive dissonance under $\tau^*$. Partially differentiating $U_i$ with respect to $\mu_i$, we get the first-order condition

$$\tau^* - \mu^* - q\gamma\left(\mu^* - \nu^*\right) = 0$$

meaning that

$$\mu^* = \frac{q\gamma\nu^* + \tau^*}{q\gamma + 1}.$$

## A1.2 Manager's Problem

We now consider the manager's problem from the standpoint of setting $w$, $\nu$, $d$, $q$ and $p$ optimally. Recall that the manager's problem is to maximize

$$V_j = 1 - w - \sigma_j\tau^* - \frac{c_m}{2}p^2 + \phi\log\left(d^*\right).$$

Substituting $\tau_i = \tau^*$ into the managerial objective function (1) and differentiating $V_j$ with respect to $w$, we obtain the following expression for the wage:[15]

$$w^* = \frac{\alpha\left(q\gamma\left(\beta_j\left(1 - \rho\nu^*\right) + \delta_\tau d^*\right)^2 + p\sigma_j^2\left(q\gamma + 1\right)^2\right) + \beta_j\rho p\sigma_j^2\left(q\gamma + 1\right)^2}{2\left(\alpha^2 q\gamma\left(q\gamma + p\right) + \alpha\beta_j q\gamma\rho\left(2q\gamma + p + 2\right) + \beta_j^2\rho^2\left(q\gamma + 1\right)^2\right)}. \quad (18)$$

In order to derive the equilibrium organizational culture $\nu^*$ and $q^*$, we take the partial derivative of $V_j$ with respect to $\nu$ and then substitute in the equilibrium wage $w = w^*$ to derive the equilibrium organizational norm:[16]

$$\nu^* = \frac{\beta_j\rho\left(\beta_j - \sigma_j\left(q\gamma + p + 1\right) + \delta_\tau d^*\right) - \alpha\sigma_j\left(q\gamma + p\right)}{\beta_j^2\rho^2}. \quad (19)$$

Likewise taking the derivative of $V_j$ with respect to $q$ and then substituting $w^*$, $\nu^*$ we see that

$$\frac{\partial V_j}{\partial q} = \frac{\alpha\gamma\sigma_j^2}{2\beta_j^2\rho^2} > 0,$$

[15] The first-order condition of the manager's problem with respect to the wage is $\frac{\sigma_j p(q\gamma+1)\left(\alpha+\beta_j\rho\right)}{\sqrt{p\left(\alpha+\beta_j\rho\right)\left(2w\left(\alpha^2 q\gamma(q\gamma+p)+\alpha\beta_j q\gamma\rho(2q\gamma+p+2)+\beta_j^2\rho^2(q\gamma+1)^2\right)-\alpha q\gamma\left(\beta_j(1-\rho\nu^*)+\delta_\tau d^*\right)^2\right)}} - 1 = 0$, from which the equation equation is derived.

[16] The first-order condition of the manager's problem with respect to $\nu$ is $-\frac{\alpha q\gamma\left(\alpha\sigma_j(q\gamma+p)+\beta_j\rho\left(\beta_j(\rho\nu-1)+\sigma_j(q\gamma+p+1)-\delta_\tau d^*\right)\right)}{\alpha^2 q\gamma(q\gamma+p)+\alpha\beta_j q\gamma\rho(2q\gamma+p+2)+\beta_j^2\rho^2(q\gamma+1)^2} = 0$, from which the equilibrium corporate culture is derived.

meaning that the manager sets the frequency of moral reminders $q^* = 1$ as high as possible regardless of where the organizational norm lies.

Similarly to before, maximizing the manager's objective with respect to monitoring probability (and again plugging $w^*$, $\nu^*$ and $q^*$ into the resulting first-order condition[17]), we obtain the equilibrium monitoring intensity:

$$p^* = \frac{\sigma_j^2 (\alpha + \beta_j \rho)}{2\beta_j^2 \rho^2 c_m}, \tag{20}$$

Finally, we need to optimize with respect to $d$:

$$\frac{\partial V_j}{\partial d} = \frac{\phi}{d} - \frac{\sigma_j \delta_\tau}{\beta_j \rho} = 0$$

giving us

$$d^* = \frac{\beta_j \rho \phi}{\sigma_j \delta_\tau}.$$

The organizational equilibrium is thus described by the equilibrium social dilemma activity equation (12), the equilibrium frequency of moral reminders ($q^* = 1$), the wage equation (18), the equilibrium corporate culture equation (19), the equilibrium monitoring intensity equation (20), and the requested performance level $d^*$. Plugging $p^*$ and $d^*$ back into the expressions for $w^*$ and $\nu^*$ we see that

$$\begin{aligned} w^* &= \frac{\alpha^2 \sigma_j^4 + 2\alpha\beta_j\sigma_j^2\rho\left(\beta_j c_m \gamma \rho + \sigma_j^2\right) + \beta_j^2 \rho^2 \sigma_j^4}{4\beta_j^4 \rho^4 c_m}, \\ \nu^* &= \frac{\phi}{\sigma_j} + \frac{1}{\rho} - \frac{\sigma_j^3 (\alpha + \beta_j \rho)^2}{2\beta_j^4 \rho^4 c_m} - \frac{\sigma_j \left(\alpha\gamma + \beta_j (1+\gamma)\rho\right)}{\beta_j^2 \rho^2} \end{aligned}$$

and

$$\tau^* = \frac{\beta_j^2 \rho^2 \left(2\beta_j^2 \sigma_j c_m \rho + 2\beta_j^2 \rho^2 c_m \phi - \sigma_j^4\right) - \alpha^2 \sigma_j^4 - 2\alpha\beta_j\sigma_j^2\rho\left(\beta_j \rho c_m \gamma + \sigma_j^2\right)}{2\beta_j^4 \rho^4 \sigma_j c_m}.$$

Meaning that the permissiveness of the organizational culture goes up when the desire to incentivize performance is greater ($\phi$ increases) but is invariant to the tradeoff between moral behavior and meeting performance targets $\delta_\tau$.

[17] The first-order condition of the manager's problem with respect to $p$ is $\frac{\sigma_j^2(\alpha+\beta_j\rho)}{2\beta_j^2\rho} - c_m p = 0$, from which the equilibrium monitoring intensity is derived.