Narratives, Organizational Ethics and Social Dysfunction

Steven Bosworth and Dennis J. Snower

31 December 2017

Abstract

This paper presents a model of the determinants of organizational culture, based on narratives and associated identities. Contractual relations between employees and managers in any organization are incomplete, and thus managers gain an incentive to use narratives to shape worker identities in order to reach the organization's goals. Different narratives give rise to different managerial moral sensitivities, inducing managers to adjust their corporate culture, in terms of the manager's efforts to instill the organizations's overt moral standard. Thereby managers affect workers' internal moral standards and consequently affect the workers' propensity to engage in socially harmful activities. Dysfunctional organizational cultures can persist in equilibrium. The model explains why cross-sectionally we see high wages in organizations with strong cultures and worker attachment, while for managers the two are substitutable incentives. We explore policy implications with an eye towards strengthening organizational cultures which discourage unethical behavior.

1 Introduction

This paper presents an analysis of organizational ethics, resulting from the social and economic interaction between managers and employees. The underlying ideas are straightforward. Many organizations face social dilemmas by engaging in activities that can generate private gains and social losses. Examples of social dilemmas that characteristically generate private gains for both employers and employees include defrauding customers, polluting the environment, and giving bribes to tax collectors; examples that generate one-sided private gains are theft by employees and employers' health and safety violations that are unknown to the employees. Such social dilemmas create tradeoffs between monetary gains and psychic losses from moral failings. This paper explores these tradeoffs and derives the implications for economic performance.

In practice, most people subscribe to a variety of inconsistent moral principles, spanning utilitarianism, deontological principles, virtue ethics. Even within these categories, people often adopt several inconsistent varieties of moral

persuasion, such as act versus rule utilitarianism; deontological principles of care, fairness, authority, etc.; and virtues associated with the diverse social roles that individuals fulfill in the course of their lives.[1] Many of these moral principles have been identified as "universal values" that are remarkably similar across cultures, but applied very differently by different cultures to concrete, practical situations.[2] The application rules for moral principles are generally provided – at least implicitly – by narratives, which give people to social roles associated with particular identities, with regard to which the moral significance of actions may be evaluated.[3]

For example, a person may adopt one identity at work (e.g. a clerk in an accounting department of a firm, doing his job amorally, i.e. following organizational requirements regardless of the social impacts of the organization) and another identity at home (e.g. a faithful and devoted husband father, acting in accordance with the moral principle of care). Generally these different identities are defined by different narratives. Some narratives make individuals sensitive to social dilemmas (leading them to attach weight to social gains and losses that are not mirrored in corresponding private gains and losses), whereas other narratives desensitize people to these dilemmas. The narrative adopted by an organization is the outcome of the narrative choices made by the employer and employees. This paper investigates the causes and consequences of these narrative choices.

The psychological mechanism whereby narratives affect people's objectives lies in a well-known insight from motivation psychology,[4] namely that all human behavior is motivated and that people have access to multiple, discrete motives, each associated with a distinct objective function. "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.[5] Motives that are associated with heightened moral sensitivities are Care (seeking to promote the well-being of others) and Affiliation (seeking belonging within social groups).[6] By contrast, motives that are linked to low moral sensitivities are Status-Seeking (seeking social standing and social influence),[7] Anger (aggressive responses to threats)[8]

---

[1]Haidt (2012) provides a very readable summary of inconsistent values relevant for the US political debate. The literature on universal values (that most people ascribe to, independently of culture, though many are inconsistent with one another) is extensive, e.g. Schwarz (2012).

[2]Fiske and Rai (2014), for example, examine differences in application rules among culturally common values.

[3]Akerlof and Snower (2016) provide a summary of various roles of narratives in economic decision making.

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

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

[6]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).

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

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

and Fear (defensive responses to threats).[9] Each of these motives is associated with a different set of preferences. Furthermore, narratives motivate: They give people social roles within a broader story, motivating them to behave in particular ways. These motives commonly have normative force.[10]

The reason why narratives are important in the economic decision making of firms and workers is that contractual relations between employers and employees are typically incomplete. Thus organizations can rely only partially on monetary incentives in order to attain managerial objectives. In addition, they adopt narratives that induce them to invest in identity formation, along lines originally proposed by Akerlof and Kranton (2005). Moving beyond the analysis of Akerlof and Kranton, we explore how the motives and identities of employers and employees, driven by underlying narratives, emerge through their workplace interactions. The resulting moral stances of employers and employees affect organizational performance.

In our analysis, narrative choices have far-reaching consequences. For the manager, narratives influence the overt moral standards of the organization, the "corporate culture" (which will be defined in terms of the frequency with which the manager reminds[11] the employees of the organization's overt moral standard), the degree of contractual completeness (which will be defined in terms of the probability that the manager observes malfeasance in the social dilemma activities) and the wage.

For workers, narratives affect their internal moral standard and the degree to which they engage in social dilemma activities. Divergences between the organization's overt moral standard and the workers' internal moral standards makes the workers suffer from cognitive and affective dissonance, to be called "manager-worker dissonance." Furthermore, divergences between the workers' internal moral standards and their social dilemma activities also lead to worker dissonance, which we shall denote as "principle-action dissonance." Since workers seek to minimize their dissonance at their workplace, the corporate culture can affect the manager-worker dissonance and thereby affect principle-action dissonance. Once these two types of dissonance become sufficiently large, the workers no longer attempt to comply with the organizational culture, thereby risking contract termination.

In this context, we explore the following causal relations linking organizational narratives to economic behaviors: (1) Different narratives give rise to different managerial moral sensitivities, whereby social losses are internalized the the manager's utility function. (2) Different managerial moral sensitives induce differences in corporate culture, in terms of the manager's efforts to instill the organizations's overt moral standard. (3) Since workers seek to avoid manager-worker dissonance, differences in corporate culture lead to differences

---

[9]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)).

[10]See Akerlof and Snower (2016).

[11]Moral reminders have been found to have a considerable influence on ethical behavior in a number of laboratory experiments. See Mazar et al. (2008); Shu et al. (2012).

in workers' internal moral standards. (4) Furthermore, since workers see to avoid principle-action dissonance, differences in the workers' internal moral standards leads to differences in the workers' social dilemma activities. (5) Finally, differences in the social dilemma activities lead to differences in marketable output and differences in social welfare.

In the organizational equilibrium, the manger's decisions concerning organizational culture and wages are consistent with the employees' decisions concerning their internal moral standard, their social dilemma activities, and their compliance with the organization's moral standards. Thus, in the organizational equilibrium, the moral standards of the manager and the employees exert mutual influence on one another. The greater the dishonorable propensities of the employees, the laxer the manager's organizational culture may become; and vice versa. In short, moral standards arise through social interactions at the workplace and these standards affect both economic and social performance. In this context, socially dysfunctional equilibria can arise in which employees perform socially undesirable levels of the social dilemma activities and managers do little to correct their behavior.

The nature of these equilibria naturally depend on the degree of contractual completeness, which we measure in terms of the probability that malfeasance is detected. In our model, this contractual completeness is endogenous: the greater the costs of detecting malfeasance and the smaller the organization's costs of generating malfeasance, the more incomplete the contracts with the employees become. The organization's costs of generating malfeasance are private costs, not public ones; consequently the equilibrium degree of contractual completeness will in general not be socially optimal.

Our analysis helps explain the wide heterogeneity across organizations in the degree of worker attachment to organizational culture – and consequently in the effectiveness of the organizational culture – that is observed in practice (Alvesson, 2002; Cha and Edmonson, 2006). Our model also explains why we observe high wages in organizations with strong cultures and worker attachment (Mühlau and Lindenberg, 2003; Masakure and Gerhardt, 2016), while for managers the two are substitutable incentives.[12] We explore policy implications with an eye towards strengthening organizational cultures which discourage unmoral behavior.

2 Model setup

We assume that a manager employs a fixed number of identical workers indexed by i. The worker engages in a "social dilemma activity" that provides both a positive private payoff and a negative social payoff. Let the private payoff per unit of worker i's social dilemma activity be normalized to unity and the

---

[12]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.

social payoff per unit of the worker's activity be Λ. Thus the marketable output produced by worker i is $Q = \tau_i$ and the resulting social welfare is $\Omega = -\Lambda\tau$.

The "social awareness" of the manager depends on the sort of narrative that drives the organization. For simplicity, let there by two narratives: N_1 that imparts high social awareness and N_2 that imparts low awareness. The manager's social awareness is denoted by the manager's disutility from worker i's social dilemma activity $\sigma_j\tau_i$, where the subscript j denotes the narrative ($j \in \{1,2\}$), $\sigma_1 > \sigma_2$. This social awareness is constrained: $\Lambda \geq \sigma_j > 0$, so that when $\sigma_j = 0$, the manager has no awareness, whereas when $\sigma_j = \Lambda$, the manager has full awareness of the social welfare implications of the social dilemma activity.[13]

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

The manager's objective function under narrative j is

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

where the payoff from the market activity is represented by unity.

Workers are assumed not to be in a position to observe directly the social harm from their activities. But they are able to observe disparities between their activities and their internal moral standards, as well as disparities between their internal moral standards and the organization's overt standard. Both of these disparities generate dissonance.

Each worker faces a tradeoff between engaging in the harmful activity τ and adhering to the particular moral standard associated with her employment. 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. This takes two forms. Each worker i maintains an internal moral standard μ_i which cannot be observed, and is asked to adhere to an organizational moral standard, $\nu^* \geq 0$, chosen costlessly be the manager . The manager chooses the frequency $q \in [0,1]$ with which the workers are reminded of the organization's moral standard ν^*. Thus q may be called the organization's "corporate culture." Workers cannot opt out of the organization's culture without running the risk of being fired.

The worker faces a straightforward tradeoff: the worker seeks the private return from the social dilemma activity τ, but this activity generates dissonance that the worker wishes to avoid. The worker faces two types of dissonance: (i) principle-action dissonance, represented by the divergence between the worker's

---

[13]We must have $\sigma_j \geq 1$ for the manager to promote ethics within the organization. If $\sigma_j < 1$, the manager may desire ceteris paribus that workers reduce activity τ, but is unwilling to pay sufficiently high wages to effect such a change.

social dilemma activity τ_i and her internal moral standard μ_i, and (ii) employer-employee dissonance, represented by the divergence between the internal moral compass μ_i and the organization's overt moral standard ν^*.

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

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

(2)

where α_j and γ are positive constants, with $\alpha_1 > \alpha_2$. Note that if the worker engages in less than the threshold level τ^* of the social dilemma activity, she earns the wage w and the private return τ_i, while paying the psychic cost for the principle-action dissonance ($\frac{\alpha_j}{2}\left(\max\{\tau_i - \mu_i, 0\}\right)^2$) and the employer-employee dissonance ($q\gamma\frac{\alpha_j}{2}\left(\max\{\mu_i - \nu^*, 0\}\right)^2$) . On the other hand, if she engages in more than the threshold level τ^* of activity τ, 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 degree of contractual completeness p, and the corporate culture q. The worker's control variables are the level of the socially harmful activity τ and the worker's moral standard μ. A switch from Narrative 1 to Narrative 2 entails a fall the manager's social awareness σ_j and a rise in the worker's sensitivity to dissonance α_j.

3 Organizational equilibrium

In the organizational equilibrium, the manager's optimal choices are consistent with the worker's optimal choices. To identify this equilibrium, we first find the contractually binding level of τ^* which is incentive-compatible. At this incentive-compatible level of τ^*, the worker is indifferent between choosing τ^* and her optimal choice of $\tau_i = \widehat{\tau}$ at which she does not comply with τ^*:

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

By the first-order condition,[14]

$$
\widehat{\tau} = \mu_i + \frac{\beta}{\alpha_j}.
$$

---

[14] $1 + \alpha(\mu_i - \widehat{\tau}) = 0$.

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

$$w + \beta\tau^* - \frac{\alpha_j}{2}\left(\tau^* - \mu_i\right)^2 - q \cdot \frac{\gamma\alpha_j}{2}\left(\mu_i - \nu^*\right)^2$$
$$= (1-p)\left(w + \beta\widehat{\tau} - \frac{\alpha_j}{2}\left(\widehat{\tau} - \mu_i\right)^2 - q \cdot \frac{\gamma\alpha_j}{2}\left(\mu_i - \nu^*\right)^2\right) + p\left(\beta\widehat{\tau} - \frac{\alpha_j}{2}\left(\widehat{\tau} - \mu_i\right)^2\right)$$

which implies the following value of the threshold level τ^*:

$$\tau^* = \mu_i + \frac{\beta}{\alpha_j} - \sqrt{\frac{p\left(2w - q\gamma\alpha_j\left(\mu_i - \nu^*\right)^2\right)}{\alpha_j}}. \tag{3}$$

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

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

which implies that that

$$\mu^* = \frac{q\gamma\nu^* + \tau^*}{1 + q\gamma} \tag{4}$$

Solving the system of indifference equations (3) and (4), we derive the incentive-compatible level of the threshold activity τ^* and the worker's internal moral standard μ^*:

$$\tau^* = \nu^* + \frac{q\gamma + 1}{q\gamma + p}\left(\frac{\beta}{\alpha_j} - \sqrt{\frac{p\left(2\alpha_j w\left(q\gamma + p\right) - \beta^2\right)}{\alpha_j^2 q\gamma}}\right) \tag{5}$$

$$\mu^* = \nu^* + \frac{1}{q\gamma + p}\left(\frac{\beta}{\alpha_j} - \sqrt{\frac{p\left(2\alpha_j w\left(q\gamma + p\right) - \beta^2\right)}{\alpha_j^2 q\gamma}}\right). \tag{6}$$

3.1 Manager's problem

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

Since the manager's utility is always decreasing in the permissiveness of the norm $(\partial V_j/\partial\nu^* = -\sigma_j)$, the norm is always set maximally: $\nu^* = 0$. Substituting $\nu^* = 0$ and $\tau_i = \tau^*$ into the managerial objective function (1) and differentiating V_j with respect to w, we obtain the following expression for the wage:[15]

$$\mathbf{w}^* = \frac{q\gamma\beta^2 + p\sigma_j^2\left(q\gamma + 1\right)^2}{2\alpha_j q\gamma\left(q\gamma + p\right)}. \tag{7}$$

---

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

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

$$q^* = \frac{\sigma_j p}{\gamma \left(\beta - \sigma_j p\right)}. \tag{8}$$

Finally, maximizing the manager's objective with respect to monitoring probability, we obtain the equilibrium degree of contractual completeness:

$$p^* = \frac{\beta}{\sqrt{2\alpha_j c_m}}, \tag{9}$$

Using the above assumptions $\alpha_j > 0$, $\gamma > 1$, and $\sigma_j \geq \beta$ it can be shown that the equilibrium corporate culture lies strictly between 0 and 1: $0 < q^* < 1$. Organizational cultures more intense than q^* will backfire in the sense that, for fixed wages, agents will revert to the behavior $\hat{\tau}$ since the incentive-compatibility condition (Eq. 3) is violated. Alternately, the organization could in theory increase w to compensate, but this would be sub-optimal in the sense that the manager's willingness to pay for ethical behavior would be less than what the agents need to identify with it. Another way to think about this is that there could be organizational cultures which agents are "not being paid enough" to identify with. Since the workers consider their outside option as entailing lower tension between their behavior and the manager's expectations, 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:

$$\frac{dw^*}{dq} = -\frac{\sigma_j^2 p \left(q\gamma + 1\right) \left(q\gamma \left(2 - p\right) + p\right) + \beta q^2 \gamma^2}{2\alpha_j q^2 \gamma \left(q\gamma + p\right)^2} < 0$$

implying that monetary incentives and organizational culture are substitutes for the manager. From the worker'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:

$$\frac{d\mu^*}{dw} = -\frac{p}{\sqrt{pq\gamma \left(2\alpha_j w \left(q\gamma + p\right) - \beta\right)}} < 0.$$

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.

---

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

4 Ethics vs. performance

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 τ, managers care about employees meeting a deliverable performance target d:

$$V_j = \sum_i \left(-w - \sigma_j \tau_i - c_m p + \theta d_i \right)$$

where the performance target measures observable worker output. The parameter θ 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 τ 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 $-\delta d_i$, with $\delta = \delta_0 - \delta_\tau \cdot \tau_i$. That is, the higher τ 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. This is the maximum performance that can be demanded from the worker.

The worker i's utility therefore becomes

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

4.1 The indifference condition

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

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

The first-order condition is $1 + \delta_\tau d^* - \alpha \left(\widehat{\tau} - \mu_i \right) = 0$, giving us

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

We can now solve the incentive compatibility condition to find τ^*:

$$w - \beta \tau^* - \frac{\alpha_j}{2} \left(\tau^* - \mu_i \right)^2 - q\gamma \cdot \frac{\alpha_j}{2} \left(\mu_i - \nu^* \right)^2 - \delta d^*$$
$$= (1-p) \left(w + \beta \widehat{\tau} - \frac{\alpha_j}{2} \left(\widehat{\tau} - \mu_i \right)^2 - q\gamma \cdot \frac{\alpha_j}{2} \left(\mu_i - \nu^* \right)^2 - \delta d^* \right) + p \left(b_j \widehat{\tau} - \frac{\alpha_j}{2} \left(\widehat{\tau} - \mu_i \right)^2 - \delta d^* \right)$$

giving us

$$\tau^* = \mu_i + \frac{\beta + \delta_\tau d^*}{\alpha_j} - \sqrt{\frac{p\left(2w - \alpha_j q\gamma\left(\mu_i - \nu^*\right)^2\right)}{\alpha_j}}. \tag{10}$$

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

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

meaning that

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

Solving the system of the indifference and first-order condition for (τ^*, μ^*) yields

$$\tau^* = \nu^* + \frac{q\gamma + 1}{q\gamma + p}\left(\frac{\beta + \delta_\tau d^*}{\alpha_j} - \sqrt{\frac{p\left(2\alpha_j w\left(q\gamma + p\right) - \left(\beta + \delta_\tau d^*\right)^2\right)}{\alpha_j^2 q\gamma}}\right) \tag{11}$$

$$\mu^* = \nu^* + \frac{1}{q\gamma + p}\left(\frac{\beta + \delta_\tau d^*}{\alpha_j} - \sqrt{\frac{p\left(2\alpha_j w\left(q\gamma + p\right) - \left(\beta + \delta_\tau d^*\right)^2\right)}{\alpha_j^2 q\gamma}}\right). \tag{12}$$

Intuitively, the worker's internalized norm and behavior are impacted both by the level of the performance target d^* and the strength of the tradeoff between performance and ethics. There is more unethical behavior and less identification with the organizational norm when performance targets increase in importance.

4.2 Manager's problem

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

$$V_j = \sum_i \left(-w - \sigma_j \tau^* - c_m p + \theta d^*\right).$$

It can still be shown that the partial derivative of V_j with respect to ν^*, $\partial V_j / \partial \nu^*$ is negative for the parameters considered. Since the manager's utility is always decreasing in the permissiveness of the norm, she sets $\nu^* = 0$.

Substituting our above expression for τ^* as well as $\nu^* = 0$ and optimizing U_m with respect to w, we obtain

$$w^* = \frac{q\gamma\left(\beta + \delta_\tau d^*\right)^2 + \sigma_j p\left(q\gamma + 1\right)^2}{2\alpha_j q\gamma\left(q\gamma + p\right)}. \tag{13}$$

We then take the partial derivative of V_j with respect to q and then substitute in the equilibrium wage $w = w^*$ to get the first-order condition for q:

$$\frac{(-q\gamma\,(\beta + \delta_\tau d^*) - q\gamma\sigma_j\,(2 - p) - \sigma_j p)\,(\sigma_j p\,(gq + 1) - q\gamma\,(\beta + \delta_\tau d^*))}{2\alpha_j q^2 \gamma\,(q\gamma + p)^2} = 0$$

giving us

$$q^* = \frac{\sigma_j p}{\gamma\,(\beta - \sigma_j p + \delta_\tau d^*)}. \tag{14}$$

We need to optimize with respect to d^*:

$$\frac{\partial V_j}{\partial d^*} = \theta - \frac{\delta_\tau \sigma_j\,(q\gamma + 1)\left(\frac{p(\beta + \delta_\tau d^*)}{\sqrt{pq\gamma\left(2\alpha_j w(q\gamma + p) - (\beta + \delta_\tau d^*)^2\right)}} + 1\right)}{\alpha_j\,(q\gamma + p)} = 0$$

giving us

$$d^* = \frac{\alpha_j \theta\,(q\gamma + p) - \beta\delta_\tau - \sigma_j\delta_\tau\,(q\gamma + 1)}{\delta_\tau^2}. \tag{15}$$

We can then substitute this back in to get an expression for q^*:

$$q^* = \frac{\sigma_j \delta_\tau}{\gamma\,(\alpha_j \theta - \sigma_j \delta_\tau)}.$$

Meaning that the strength of the organizational culture gets weaker when the desire to incentivize performance is greater (θ increases) but stronger when the tradeoff between moral behavior and meeting performance targets becomes steeper (δ_τ increases).

Finally, maximizing with respect to p yields

$$p^* = \frac{\sigma_j \delta_\tau\left(\sqrt{2}\alpha_j \theta + 2\delta_\tau \sqrt{\alpha_j c_m}\right)}{(\alpha_j \theta - \sigma_j \delta_\tau)\left(\sqrt{2}\alpha_j \theta - 2\delta_\tau \sqrt{\alpha_j c_m}\right)}. \tag{16}$$

Intuitively, the more important performance targets are to the manager, the laxer the organizational culture becomes. Though the manager is able to use the organizational culture to induce higher ethics for a given level of performance target, this creates cognitive dissonance among the workers and ultimately more performance-oriented firms have worse ethics. Interestingly, in organizations where there is a very strong tradeoff between performance and ethics (i.e. workers very tempted to compromise ethics to meet performance targets), managers implement stronger organizational cultures. The implications of these results are important when considering the design of organizational incentives. While increasing the importance of incentives may suggest that maintaining a strong ethical culture as a countervailing weight becomes critical, it is not in managers' interest to do so. If the consequences of a lax organizational culture spill over into other domains, then perverse incentives can have even worse implications than neoclassical economics would suggest.

5 Narrative shifts

Narratives coordinate the degree to which both workers and managers care about the workers' ethical behavior. In the model, these are conceived of as α_j and σ_j, respectively. While managers would benefit from workers who care more about ethics (higher α_j), such a narrative shift would likely also entail the manager herself being more concerned with ethics (higher σ_j). The manager's utility is however decreasing in σ_j, since this means she feels the ethical lapses of the workers more keenly. Having wide latitude to affect the organizational narrative then, the manager needs to consider whether she can get workers to care sufficiently more about ethics to also be willing to care more about ethics herself. In terms of the model parameters, when will managers pursue narrative shifts which increase α_j at the expense of increasing σ_j? On the one hand, the firm would be able to demand higher ethical standards from the workers, on the other, the manager would herself need to take ethical breaches much more seriously. Suppose that a shift from norm $N_1 = \{\alpha_1, \sigma_1\}$ to norm $N_2 = \{\alpha_2, \sigma_2\}$ with $d\alpha_j \approx \alpha_2 - \alpha_1 > 0$ and $d\sigma_j \approx \sigma_2 - \sigma_1 > 0$ entails shifts $d\alpha_j$ and $d\sigma_j$ in the parameters α and σ, respectively. To understand whether the manager prefers such a shift we can apply the envelope theorem to V_j:

$$dV_j = \frac{\partial V_j}{\partial \alpha_j} d\alpha_j + \frac{\partial V_j}{\partial \sigma_j} d\sigma_j$$

where the manager requires $dV_j > 0$. This means:

$$\frac{d\alpha_j}{d\sigma_j} \cdot \frac{\sigma_j}{\alpha_j} > \frac{q\gamma \left(\beta + \delta_\tau d^*\right) - \sqrt{pq\gamma \left(2\alpha_j w \left(q\gamma + p\right) - \left(\beta + \delta_\tau d^*\right)^2\right)}}{q\gamma \left(\frac{p\left(\left(\beta + \delta_\tau d^*\right)^2 - \alpha_j w \left(q\gamma + p\right)\right)}{\sqrt{pq\gamma \left(2\alpha_j w \left(q\gamma + p\right) - \left(\beta + \delta_\tau d^*\right)^2\right)}} + \beta + \delta_\tau d^*\right)}. \tag{17}$$

In terms of the parameters, this simplifies to

$$\frac{d\alpha_j}{d\sigma_j} \cdot \frac{\sigma_j}{\alpha_j} > \frac{2 \left(4\alpha_j^{5/2} c_m^{3/2} \delta_\tau^3 - 2\delta_\tau \theta^2 \sqrt{\alpha_j^7 c_m} + \sqrt{2}\alpha^4 \theta^3 - 2\sqrt{2}\alpha_j^3 c_m \delta_\tau^2 \theta\right)}{\alpha_j^2 \left(\sqrt{2}\alpha_j \theta - 4\delta_\tau \sqrt{\alpha_j c_m}\right) \left(\alpha_j \theta^2 - 2c_m \delta_\tau^2\right)}.$$

I.e. the associated percentage increase in α relative to the percentage increase in σ must be sufficiently large.

Conversely, the narrative may shift in a direction more conducive to laxer ethical standards. If the manager perceives that she can herself care less about ethical breaches without adversely affecting workers' concerns too much, then she may shift the narrative in a direction which is more conducive to bad behavior.

6 Discussion

Transforming organizational cultures is a matter of crucial importance in today's society. Many workplaces for example are grappling with cultures of sexual ha-

rassment and taking long-overdue steps to root out instances of abuse. In terms of our framework, one can conceptualize this as outside pressure reducing the monitoring costs c_m. 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] Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. Educational Psychology Review, 13(2), 73-92.

[6] Fiske, A. P. & Rai, T. S. (2014). Virtuous Violence, Cambridge, UK: Cambridge University Press.

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

[8] Haidt, J. (2012). The Righteous Mind. London: Penguin.

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