Why do Firms have 'Purpose'? The Firm's Role as a Carrier of Identity and Reputation: Online Appendix

Rebecca Henderson and Eric Van den Steen[*]

January 12, 2015

1 Introduction

This is the Online Appendix to "Why do Firms have 'Purpose'? The Firm's Role as a Carrier of Identity and Reputation." It describes the model in more detail and provides the formal analysis.

2 Model

Consider a setting with N firms with $N \to \infty$, each having 1 manager (M) and K employees (E). (With less than K employees, the firm produces nothing.) Managers are exogenously attached to firms prior to the start of the game, with exactly one manager per firm. Employees come from an infinite labor pool of potential employees, i.e., more than sufficient employees for all firms. Employees will join firms as part of the game, following a hiring process in the form of a cooperative game using the core solution. (This produces an efficient matching with wages to support it.) If there is more than one core solution, then one gets selected at random with all being equally likely.

Managers and employees come in two types: Social (S) or Asocial (A). In the pool of potential employees, there is a finite number L of Social types, with $K < L$, and thus an infinite number of Asocial types. It is, for simplicity, also common knowledge that exactly one manager is of type S (though it is not publicly known which one).

[*] Henderson: Harvard Business School, Boston, MA 02163, rhenderson@hbs.edu. Van den Steen: Harvard Business School, Boston, MA 02163, evandensteen@hbs.edu. We thank Bob Gibbons both for organizing the session and for many inspiring discussions, and the AEA participants for helpful comments.

Players' types are, at the start of the game, unknown to the player and to anyone else. A player 'realizes' his or her type when called upon to make a payoff-relevant choice (which fits the idea of 'utility' as an as-if construct based on revealed preference). However, players may 'forget' their type and will then need to infer it from their earlier actions. Employees also observe the manager's type when considering a wage offer or making other payoff-relevant choices. Again, they may forget that information.[1] This 'forgetting' captures the fact, well-supported by psychological evidence, that people often don't 'know' or 'remember' their preferences but infer these from their own actions. While this may be driven by literal forgetting, our interpretation is that it captures the fact that some things are more focal than others – for example, because we get reminded of them on a regular basis – and are thus more likely to affect our thinking and our 'utility'. Formally, in terms of what players forget and remember, we will assume that the game is almost completely standard, with one exception: it is common knowledge that in the last period (when utilities are realized), with probability q, players will remember only publicly observable and/or commmon knowledge facts (incl. the nature of the equilibrium), and will not 'remember' anything else. In that event, all that anyone – incl. the employee – knows is: in which firm a particular employee works, what choices that firm made, and what the equilibrium of the game is.

In the course of the game, firms will choose an action Z (through their employees and/or their manager) from a set of four potential actions $\{X_h, X_l, Y_l, Y_h\}$ with respective monetary payoffs $\mathcal{P}_Z$ being $P > P - \epsilon > -P + \epsilon > -P$ for some $P >> \epsilon > 0$. Absent other considerations, everyone would thus always want to choose X_h. But actions also have a social impact $\mathcal{B}_Z$, which S-type players care about (and which is the focus of a company's potential purpose). The social impact of the four alternatives $\{X_h, X_l, Y_l, Y_h\}$ are a permutation of $\{B, B, 0, 0\}$, for $B > 0$.

In order to explore the effect of different correlations between monetary payoffs and social benefits, we consider three cases in the analysis:

- *Positive* correlation: the social impact of the actions $\{X_h, X_l, Y_l, Y_h\}$ are $\{B, B, 0, 0\}$. This is an extreme 'doing well by doing good' case: the socially most helpful choices are also the most profitable ones.

[1] Managers' reputations can be endogenized by letting managers take costly actions that may signal their type. Modeling this explicitly would add complexity to the analysis without any clear benefit.

- *No* correlation, which is the most important case from a practical perspective: the social benefits are a permutation of $\{B, B, 0, 0\}$ with each permutation equally likely. In this case, there is essentially no ex-ante relationship between social effect and payoff of a choice.

- *Negative* correlation: the social benefits of the actions $\{X_h, X_l, Y_l, Y_h\}$ are $\{0, 0, B, B\}$, i.e., the socially most helpful actions are the least profitable (or most unprofitable).

The social effect of all choices will be observed publicly during the game (incl. in the last period).

The action is chosen either by a randomly selected employee (each with probability ρ) or by the manager (with complementary probability $1 - K\rho \geq 0$). If an employee chooses the action, then the manager can overturn it at a cost $\delta \downarrow 0$ to both the manager and the employee. In that case, the manager can make the new choice. (The manager and the selected employee get the private social benefits from the action that is actually executed.)

When selected to make the firm's choice, an employee also needs to decide whether or not to exert effort. Exerting effort carries a personal cost c. Employees' effort determines the likelihood ψ that the action gets executed. Let e_k be the indicator that the selected employee k exerts effort, then the probability of execution $\psi = \underline{\psi} + \delta e_k$ for $\underline{\psi} \in (0, 1)$, $\delta \in (0, (1 - \underline{\psi})]$. In other words, when an employee decides to exert effort, the probability of execution increases with discrete amount δ; in all other cases (incl. when the manager makes the choice), the probability of execution is $\underline{\psi}$. If the firm's action gets executed, then its payoff is as discussed above. If it does not get executed, then the firm has no payoff and the public cannot observe which choices the firm made.

Utilities Whereas A-type players don't care about the social impact, S-type players who are personally involved in the choice get a private benefit $\gamma_i \mathcal{B}_Z$, with $i \in \{M, E\}$, $\gamma_M > 0$, and $\gamma_E \geq 0$. We will say that an employee is personally involved when she is selected to make the choice, whereas the manager is always personally involved. (Note that an employee who was selected to make the choice will get the private benefit of the final choice, even if she originally chose something different but was overturned. The interpretation is that it is still this employee who executes the action and thus feels the private benefits (or costs) from doing so. The manager gets the private benefit even when she does not overturn. The interpretation is that she still enables or allows the action, even if she does not personally choose it, and thus experiences the private benefits. Alternative assumptions are possible and seem to give similar results.) Note that the private benefits will thus be a permuation of $\{\gamma_i B, \gamma_i B, 0, 0\}$ with $i \in \{E, M\}$. We will look explicitly at cases where $\gamma_E = 0$, as we are interested in settings where the social effect of the employees' own actions is very modest.

In terms of the monetary payoffs (or incentives), each manager will get a share α_M of her firm's overall profit Π (consisting of payoff $\mathcal{P}_Z$ minus employee wages) plus a wage that we normalize for simplicity to $w_M = 0$. Employees get a share α_E of monetary payoffs $\mathcal{P}_Z$ plus a wage w_E that is determined in the matching process, as part of the game.[2]

For managers' utilities, this implies the following:

- A manager of type A ('Asocial') only cares about profits. In particular, an A-manager has a utility $\alpha_M \Pi$ where Π is her firm's profit (incl. employees' wages) prior to the manager's pay.

- A manager of type S ('Social') cares about both profit and the social impact of her – i.e., her firm's – actions: $\gamma_M \mathcal{B}_Z + \alpha_M \Pi$.

Employees care not only about the action's direct (social and profit) implications but also about their identity and reputation, i.e., about their own beliefs about themselves and about others' beliefs about them. In particular, employees get some benefit from a 'Social' identity, i.e., from holding a self-belief that they are of type S, and from a 'Social' reputation, i.e., from others' believing that they are of type S. (This reflects the central assumption that S-type actions are socially valued, i.e., that the firm's purpose reflects a social value.) Let μ denote an employee's belief about herself that her type is S and let ν denote the outsiders' beliefs (based on what these outsiders have observed) about the employee's type. (Any particular outsider will observe only one firm at a time. In other words, outsiders cannot use the behavior of the full population of firms to form their beliefs.) The employee's utility will

[2] As with other elements of the model, the specific assumptions are chosen to simplify the analysis. In this case, for example, giving the employee $\alpha_E \Pi$ or the manager $\alpha_M \mathcal{P}_Z$ gives qualitatively similar results but with more complex analysis. Take, for example, the case of giving the employee $\alpha_E \Pi$. As Π depends on the employees' wages, which depend on their turn on the employees' expected utility, the problem would require a fixed point calculation, which makes things a lot more complicated.

then have a term $U(\mu, \nu)$ which is increasing in both its arguments. We will immediately be more specific about this utility from identity and reputation, but let me first specify now the employees' utilities.

- An employee of type A only cares about monetary payoffs and about reputation/identity: $\alpha_E \mathcal{P}_Z + U(\mu, \nu)$.
- An employee of type S also cares about the social impact of her action:[3] $\gamma_E \mathcal{B}_Z + \alpha_E \mathcal{P}_Z + U(\mu, \nu)$ where $\gamma_E \mathcal{B}_Z$ is the employee's private benefit from the firm's action when she was selected to choose that action.

Note that we make here the simplifying assumption that managers don't care about identity or reputation. While the assumption is made pureful for analytical convenience, it can be motivated by the fact that, given their visibility, they have other sources for reputation and identity. We will discuss below how the results extend when changing this assumption.

We assume that the employee's utility from identity and reputation $U(\mu, \nu)$ takes the form $U = \lambda(\mu+\nu)+(1-\lambda)\mu\nu$. The purpose of this parametrization is to explore, through λ, the effect of identity and reputation being complements. Note also that this functional form implies the normalization that $U(0, 0) = 0$.

Timing The timing is then as follows, as also captured in Figure 1:

1. Hiring
   (a) Through a cooperative hiring game, the potential employees get allocated to firms and their wages set – with non-matched potential employees receiving the outside option $w = 0$ (plus any utility U from reputation and identity). In particular, the allocation of employees and wages must be a core allocation, given the continuation equilibrium. (If more than one core allocation exists, then one is selected at random with all being equally likely. This only matters for the allocation of S-employees to A-firms.) However, to capture the fact that managers cannot observe potential employees' types, employee wages in a firm will be constrained to

[3] It would seem reasonable to assume that an S-employee also cares more about social reputation and identity. Interestingly, this result emerges endogenously when identity and reputation are complements. We use the latter to capture the former through micro-foundations.

be identical.[4] (In particular, the firm cannot 'wage-discriminate' between S- and A-employees.)

2. Action Choices

   (a) In each firm, one player is selected to take actions. (Each of the employees is selected with small probability ρ. The manager is selected with complementary probability $1 - K\rho$.)

   (b) If the selected player is an employee, she decides whether to exert effort at personal cost c.

   (c) The social impact of the four alternatives are drawn and publicly revealed. The selected player chooses an action from $\{X_h, X_l, Y_l, Y_h\}$.

   (d) The manager can overturn the action chosen. Overturning costs $\delta \downarrow 0$ to both the manager and to the selected employee/player. When an action is overturned, the manager chooses which action is implemented.

   (e) With probability ψ, the chosen action is executed. (With complementary probability, the firm has no action.)

3. Payoffs

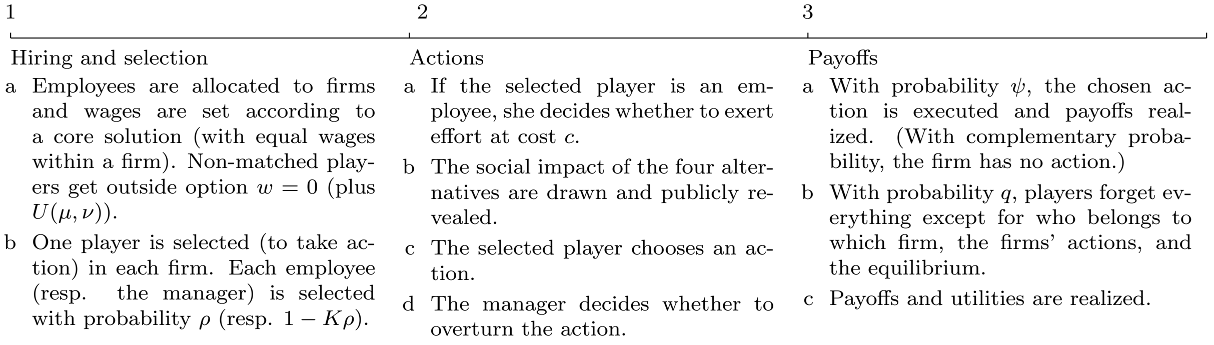
   (a) With probability q, players forget everything except for who belongs to which firm, the different firms' actions (and their monetary payoffs and social impact), and the equilibrium that is being played.

   (b) Payoffs are realized.

Neither firms nor managers have public labels, and they can therefore not be distinguished by the public, except through the actions they took.

We will also make a few parametric assumptions to focus on the most natural setting for firms. First, we assume that $\alpha_M > \alpha_E$, i.e., managers care relatively more about firm performance than employees. This is a very natural assumption as managers typically have more firm-wide incentive pay and as their fate is typically more closely tied to the performance of their firm. We will also assume that the $\alpha_E \mathcal{P}_Z$ come out of the firm's revenue so that the firm is left with $\Pi = (1 - K\alpha_E)\mathcal{P}_Z + Kw_E$. The manager's pay

[4] This may be derived endogenously, but imposing it makes it much simpler.



1

Hiring and selection

a Employees are allocated to firms and wages are set according to a core solution (with equal wages within a firm). Non-matched players get outside option $w = 0$ (plus $U(\mu, \nu)$).

b One player is selected (to take action) in each firm. Each employee (resp. the manager) is selected with probability ρ (resp. $1 - K\rho$).

2

Actions

a If the selected player is an employee, she decides whether to exert effort at cost c.

b The social impact of the four alternatives are drawn and publicly revealed.

c The selected player chooses an action.

d The manager decides whether to overturn the action.

3

Payoffs

a With probability ψ, the chosen action is executed and payoffs realized. (With complementary probability, the firm has no action.)

b With probability q, players forget everything except for who belongs to which firm, the firms' actions, and the equilibrium.

c Payoffs and utilities are realized.

Figure 1: Timing

comes from that profit, so that the firm's ultimate profit is $(1 - \alpha_M)\Pi$ (given that $w_M = 0$).

We assume further that $P - \epsilon/2 > \max\{\gamma_M B, \gamma_E B, KU(1,1)\}$, i.e., that the monetary payoff effect of the firm's action is larger than the private benefits or identity effects that it generates. This assumption on P only plays a role in the case of negative correlation. The motivation for these assumptions is that we are mainly interested in for-profit firms and our conjecture is that firms that do not satisfy these conditions – i.e., where the 'social' effects dominate profits – will be organized as non-profits. Combining for-profits and non-profits in one analysis can be confusing in this case. (This does generate an interesting research question, however: whether this conjecture about non-profits is correct and, if so, why.)

3 Analysis

Let 'S-firm' denote the firm with the S-manager and 'A-firms' the firms with A-managers. We will disregard – both in the statement of the proposition and in the proof – those settings with exact equalities in parameters as these create a lot of additional complexity at very limited gains.

The following proposition then captures the results.

- The first part of the propsosition shows that for the social purpose to have any effect, there must be a trade-off between profits and social effects: firms are indistinguishable and there is no social reputation/identity when $\mathcal{P}_Z$ is perfectly correlated with $\mathcal{B}_Z$. It also shows that the credibility of purpose depends on the manager's values: if γ_M is too small, then the firm will end up simply taking the most profitable action and there is no social identity or reputation.

- The second part of the proposition shows that, under the right conditions, social purpose can lead to social identity and reputation and it can do so profitably. This effect requires at least some trade-off between profits and social impact and a manager who cares sufficiently to effectively make that trade-off. (This can happen even when B, ρ, and/or γ_E are small: even a small social impact or private benefit can result in sorting, which then leads to a beneficial reputation and identity effect.) In that case, the employees can get a social identity and reputation. In exchange, employees accept lower wages. Moreover, they will also exert more effort, as their positive reputation and identity depend on the firm's action being executed. As all of this does not reduce the profits of the A-firms, industry profits will be higher. Moreover, the increased effort can even make the social impact higher than in the case with perfect correlation. The profit difference increases in the number of employees, which reflects the scale benefits, and on the likelihood of being selected (as that gives employees more direct private benefits from the social actions).

- The third part shows that negative correlation does not necessarily imply that purpose comes at lower profitability: the increase in effort can more than compensate for the decrease in monetary payoffs. In other cases, however, that is not the case and then purpose leads to lower profits.

- The proposition also points to the importance of the complementarity of identity and reputation. It shows that, *when there is no complementarity between identity and reputation*, at $\rho = 0$ (or $\gamma_E B = 0$), there is no sorting of S-employees to S-firms because all employees get the same differential utility from being at an S-firm. While the $\rho = 0$ case is extreme, it does indicate that the complementarity will lead to much stronger sorting. The complementarity has the effect that S-employees care endogenously more about S-reputation and -identity than A-employees.

Proposition 1 • *The S-firm is for an outsider indistinguishable from* *the A-firms – with the same wages, expected effort, actions, and expected profits, and with the same social reputation – in the case with* positive *correlation (between $\mathcal{P}_Z$ and $\mathcal{B}_Z$), and in the cases with sufficiently low $\gamma_M B$.*

- *The S-firm can have higher profits than the A-firms (even when $\rho\gamma_E B = 0$), both in the case with no correlation and in the case with negative correlation and effort. In such cases, S-employees have a social reputation and identity, accept lower wages, and exert more effort. Industry profits are higher and social impact can also be higher than in the above case with indistinguishable firms. The profit difference increases in firm size K and employee involvement ρ. With no correlation or effort, profits are highest at intermediate levels of $\gamma_M B$.*
- *The S-firm has lower profits than the A-firms in some (but not all) cases with* negative *correlation (in particular when $\gamma_M B > \alpha_M(2P - \epsilon)$ and there is no effort) – when S-employees' lower wages do not make up for the lower monetary payoffs.*
- *Whenever $\lambda = 1$ ('no complementarity') and $\rho\gamma_E B = 0$, the S-firm has lower profits than the A-firms. In such cases, the S-firm chooses the social action, but its employees are randomly drawn, have* no *social reputation, and are paid the same as A-firm employees.*

The last case is the one case where the firm's pro-social actions are purely a private consumption by the manager with no benefit at all to the firm. The reason is that in this case, there is no sorting of employees.
Proof : Note first the following:

- The outside options for A- and S-type potential employees are respectively 0 and $(1-q)U(1,0)$.
- When the action fails to execute, outsiders' beliefs are $\mu = \nu = 0$. The reason is that 1) firms that take no actions are indistinguishable by outsiders and 2) in every firm (with $N \to \infty$ firms), actions fail to execute with (at least) positive probability $1-(\underline{\psi}+K\rho\delta)$. So there is an infinite number of fails by firms with almost surely no S-employees. As $\mu = \nu = 0$ in such case, it further follows that an employee's expected utility when the action fails to execute is w for an A-type employee and $w + (1-q)U(1,0)$ for an S-type employee.

- By the time that the manager chooses (or overturns) actions, the wages and effort choices are fixed. Moreover, the manager's choice will only have an effect if it gets executed. So, when making her choice, the manager will condition on her action getting executed and will disregard any effect on wages or effort.

- At the time that employees make their effort choices, wages are fixed but actions are not yet chosen. Employees will thus exert effort if the increase if their expected continuation utility exceeds the cost of effort. Let e_k be the indicator that employee k exerts effort (in equilibrium) when selected. Let K denote a firm's full set of employees and K_{-k} that set excluding employee k.

Consider then first an A-firm, i.e., a firm with an A-manager. As an A-manager only cares about profits and as $\delta \downarrow 0$, the A-manager will overturn any choice other than X_h and turn it into X_h. Employees of A-firms will therefore always choose X_h: they anticipate getting overturned (at cost $\delta \downarrow 0$ to themselves) if they choose an action different from X_h and will anyways get only the private benefit from the ultimate action (i.e., X_h), independent of what they chose themselves first. It thus also follows that the ultimate choice of A-firms will always be X_h.

Consider next the one S-firm, i.e., the one firm with an S-manager. Note that the S-manager will prefer an action with higher $\gamma_M \mathcal{B}_Z$ over an action with higher profits if the gain in private benefit outweighs α_M times the loss in profit.

In the perfect correlation case, the S-manager always prefers X_h and will overturn any other choice (into X_h). Following an argument analogous to above, employees in an S-firm will therefore always choose X_h; the manager will overturn anything else; the firm will always end up choosing X_h; and A-firms and S-firms are indistinguishable.

In the two other correlation cases, the manager trades off $\gamma_M \mathcal{B}_Z$ against $\alpha_M \mathcal{P}_Z$ (given that the wages and effort parts of Π are already fixed, leaving only $\mathcal{P}_Z$ and social benefits). (For the argument below, note that the cases that need to be considered are limited by the following fact: either at least one of $\{X_h, X_l\}$ has $\mathcal{B}_Z = B$ or both Y_h and Y_l have $\mathcal{B}_Z = B$.)

- If $\gamma_M B < \alpha_M \epsilon$, then an S-manager will always prefer X_h and overturn anything else (as $\delta \downarrow 0$). Per the earlier argument, employees will also always choose X_h. In this case, a S-firm thus acts completely like an A-firm, and the two are again indistinguishable.

- If $\alpha_M \epsilon < \gamma_M B < \alpha_M(2P - \epsilon)$, an S-manager prefers X_l with $\mathcal{B}_Z = B$ over X_h with $\mathcal{B}_Z = 0$, but prefers X_h with $\mathcal{B}_Z = 0$ over Y_l with $\mathcal{B}_Z = B$.
- If $\alpha_M(2P - \epsilon) < \gamma_M B$, then an S-manager prefers Y_l with $\mathcal{B}_Z = B$ over X_h with $\mathcal{B}_Z = 0$ and will thus always choose the action with $\mathcal{B}_Z = B$ with the highest $\mathcal{P}_Z$, which can be any action from $\{X_h, X_l, Y_l\}$.

We can now determine the full equilibria. The rest of the proof will for every case first consider the case without effort – i.e., $c = \infty$ – and then consider how things change with effort. Throughout, when talking about firm profits, we will consider profits before the manager's share (α_M). This is just a scaling that saves on notation and does not affect any conclusions as it applies to all profits. Unless otherwise noted, we will assume in what follows that either $\lambda < 1$ or $\rho \gamma_E B > 0$.

Perfect correlation Consider first the case with perfect correlation and no effort. In that case, employees in both A- and S-firms always choose X_h with $\mathcal{B}_Z = B$. As firms are indistinguishable by outsiders, any outsider will hold the same belief for all employees, which must be $\nu = 0$. (Note that as $N \to \infty$, there are an infinite number of employees of which only a finite number (L) can be of type S. As $N \to \infty$, $\nu \downarrow 0$.) Moreover, upon forgetting her own type, every employee will also believe that $\mu = 0$. It follows that every particular employee gets the same payoff from any firm: an A-employee gets $w + \underline{\psi} \alpha_E P$ and an S-employee gets $w + \underline{\psi}(\alpha_E P + \rho \gamma_E B) + (1-q)U(1,0)$. Their outside options are respectively 0 and $(1-q)U(1,0)$. It is then straightforward to show that in all core allocations (with equal wages within a firm), all employees get a wage $w = -\underline{\psi} \alpha_E P$; all S-employees are employed; the S-employees are randomly distributed among the different firms. It follows that firms in the perfect correlation case all have profit $\underline{\psi} P$ (prior to the manager's pay).

Consider now what changes with the addition of effort. As the expected utility conditional on execution is as above, A-employees and S-employees will exert effort iff respectively $\delta \alpha_E P > c$ and $\delta(\alpha_E P + \gamma_E B) > c$ (where we have conditioned – for the S-employee – on being selected, as she only then can exert effort). S-employees thus exert more effort than A-employees.

Their respective expected utilities increase to $w + (\underline{\psi} + \sum_{i \in K_{-k}} \rho\delta e_i)\alpha_E P + \rho(\delta\alpha_E P - c)e_k$ and $w + (\underline{\psi} + \sum_{i \in K_{-k}} \rho\delta e_i)\alpha_E P + \rho(\delta(\alpha_E P + \gamma_E B) - c)e_k + (1-q)U(1,0)$. It follows that the difference in expected utility between an S-employee and an A-employee is always the same (positive) amount, for any firm (or for any other employees) with which the employee is matched. It follows that in all core allocations, all S-employees will be employed but will be randomly distributed among the different firms.[5] Moreover, as $N \to \infty$ and the number of S-employees is $L < \infty$, the likelihood of one or more S-employees in any particular firm converges to zero. It is then straightforward to show that in all core allocations (with equal wages within a firm), all employees get a wage $w = -\underline{\psi}\alpha_E P$ if $\delta\alpha_E P < c$ and $w = -(\underline{\psi} + K\rho\delta)\alpha_E P + \rho c$ if $\delta\alpha_E P > c$. It follows that firms in the perfect correlation case all have profit $\underline{\psi}P$ if $\delta\alpha_E P < c$ and $(\underline{\psi} + K\rho\delta)P - K\rho c$ if $\delta\alpha_E P > c$. Overall, whereas S-employees exert more effort, firms still have the same actions, wages, and expected profits.

Negative correlation Consider next the case with negative correlation. Assume first that $\gamma_M B < \alpha_M(2P - \epsilon)$ and consider the case with no effort. In that case, employees in both A- and S-firms always choose X_h with $\mathcal{B}_Z = 0$. As firms are again indistinguishable by outsiders, any outsider will hold the belief that $\nu = 0$, while employees will also believe that $\mu = 0$ upon forgetting their own type. It follows that any particular employee gets the same payoff from any firm: an A-employee gets $w + \underline{\psi}\alpha_E P$ and an S-employee gets $w + \underline{\psi}\alpha_E P + (1-q)U(1,0)$. Their outside options are again respectively 0 and $(1-q)U(1,0)$. It is then straightforward to show that in all core allocations (with equal wages within a firm), all employees get a wage $w = -\underline{\psi}\alpha_E P$; a random selection of employees (both A-types and S-types) are employed and randomly distributed across firms; and firms all have profit $\underline{\psi}P$.

In the case with effort and $\gamma_M B < \alpha_M(2P - \epsilon)$, firms are still indistinguishable. Moreover, conditional on being selected, S-employees will exert the same effort as A-employees. A random selection of employees (both A-types and S-types) will thus again be employed and randomly distributed across

[5] It may, at first, seem that it would be efficient to sort S-emmployees together as their utility is higher and they exert more effort, which scales up their utility. Note, however, that the effort of an employee increases everyone's payoff from $\mathcal{P}_Z$ but only that employee's payoff from $\mathcal{B}_Z$. So it scales up only the common part of other employees. Hence, there is no efficiency gain from sorting.

firms in any core allocation. Qualitatively, the equilibrium thus remains unchanged with all firms having the same actions, wages, and expected profits, and with a random selection of employees. (The wages are now $w = -\underline{\psi}\alpha_E P$ if $\delta\alpha_E P < c$ and $w = -(\underline{\psi} + K\rho\delta)\alpha_E P + \rho c$ if $\delta\alpha_E P > c$; the firms' profits change analogously.)

Consider next the case with negative correlation, $\gamma_M B > \alpha_M(2P - \epsilon)$, and no effort. In this case, A-firms/managers – and thus employees of A-firms – will all choose X_h (with $\mathcal{B}_Z = 0$), whereas S-firms/managers – and thus employees of S-firms – will all choose Y_l (with $\mathcal{B}_Z = B$). An A-employee gets payoff $w + \underline{\psi}\alpha_E P$ from an A-firm and $w + \underline{\psi}[\alpha_E(-P+\epsilon) + qU(\mu,\nu) + (1-q)U(0,\nu)]$ from an S-firm. An S-employee gets payoff $w + \underline{\psi}\alpha_E P + (1-q)U(1,0)$ from an A-firm and $w + \underline{\psi}[\alpha_E(-P+\epsilon) + \rho\gamma_E B + qU(\mu,\nu) + (1-q)U(1,\nu)] + (1-\underline{\psi})(1-q)U(1,0)$ from an S-firm. The efficient allocation of employees is then that the S-firm employs S-employees, whereas A-firms employ a random selection of employees. The public will then infer that $\nu = 1$ for the S-firm and $\nu = 0$ for the A-firms. Similarly, upon forgetting, an employee of an S-firm will infer that $\mu = 1$ whereas an employee of an A-firm infers that $\mu = 0$. (And these beliefs are obviously correct in expectation.) The core allocation is then that the A-firms employ a random selection of employees at wage $w = -\underline{\psi}\alpha_E P$, whereas the S-firm employs S-employees at wage $w = -\underline{\psi}[\alpha_E(-P+\epsilon) + \rho\gamma_E B + [U(1,1) - (1-q)U(1,0)]]$. The profits of A-firms (prior to the manager's pay) equal $\underline{\psi}P$, whereas the profit of the S-firm equals $\underline{\psi}[(-P+\epsilon) + K\rho\gamma_E B + K(U(1,1) - (1-q)U(1,0))]$. In this case, the S-firm thus has lower profits than A-firms as we assumed that $K\rho\gamma_E B + K(U(1,1) - (1-q)U(1,0)) < 2P - \epsilon$.

Consider now effort in this case with $\gamma_M B > \alpha_M(2P - \epsilon)$. Employees (of both A- and S-type) will exert effort in an A-firm iff $\delta\alpha_E P > c$ (as usual) whereas an S-employee in an S-firm will exert effort iff $\delta(\alpha_E(-P+\epsilon) + \gamma_E B + qU(\mu,\nu) + (1-q)U(1,\nu) - (1-q)U(1,0)) > c$. This may now result, however, in higher profits for the S-firm than for the A-firms, for example when $\underline{\psi} = 0$ and $\delta\gamma_E B > c > \delta\alpha_E P$.

Consider now also the special case (for $\gamma_M B > \alpha_M(2P - \epsilon)$) with $U = \mu + \nu$ (i.e., $\lambda = 1$), $\rho\gamma_E B = 0$, and no effort. In this case, an A-employee gets payoff $w + \underline{\psi}\alpha_E P$ from an A-firm and $w + \underline{\psi}[\alpha_E(-P+\epsilon) + q(\mu+\nu) + (1-q)\nu]$ or $w + \underline{\psi}[\alpha_E(-P+\epsilon) + q\mu + \nu]$ from an S-firm. An S-employee gets payoff $w + \underline{\psi}\alpha_E P + (1-q)$ from an A-firm and $w + \underline{\psi}[\alpha_E(-P+\epsilon) + q(\mu+\nu) + (1-q)(1+\nu)] + (1-\underline{\psi})(1-q)$ or $w + \underline{\psi}[\alpha_E(-P+\epsilon) + q\mu + \nu] + (1-q)$

from an S-firm. It follows that the S-employee always gets $1 - q$ more than the A-employee, independent of which firm (or which other employees) the S-employee is allocated to, so that allocating S-employees to S-firms is now no more efficient than allocating them to A-firms. It further follows that any allocation can be in the core and that there is no sorting in equilibrium. But that further implies that $\mu = \nu = 0$ and thus the payoffs for the S-firm become lower for both types of employees than for the A-firm. This finally implies that the S-firm is less profitable than the A-firms. And, as before, considering also effort preserves these results as the net utility of A- and S-types is the same for the same firm, so that they exert the same effort in the same firm (but more effort in A-firms than in S-firms, accentuating the profit difference).

No correlation

Case $\gamma_M B < \alpha_M \epsilon$ Consider first the case with $\gamma_M B < \alpha_M \epsilon$ and no effort. In that case, S-managers behave just like A-managers, never considering social impact and always choosing X_h, which results in an expected social impact of $B/2$. All firms look alike to all employees; A- and S-employees have respective payoffs $w + \underline{\psi}\alpha_E P$ and $w + \underline{\psi}[\alpha_E P + \rho\gamma_E B/2]$ from any firm. It follows that in any core allocation, all potential S-employees are employed and randomly distributed over firms; all employees receive a wage of $w = -\underline{\psi}\alpha_E P$; and all firms earn $\underline{\psi}P$.

In the case with effort, S-employees will exert more effort than A-employees, but to the same degree in any firm they are employed at. The core allocation is thus still that all potential S-employees are employed and randomly distributed over firms. Ex-ante – given that $L < \infty$ and $N \to \infty$ – firms will thus still be indistinguishable with identical wages, expected effort, actions, and expected profits. Moreover, as effort itself is never observed – only whether actions execute or not – firms with higher effort will also be indistinguishable from firms with luck, making again all firms with actions indistinguishable. Hence, A- and S-firms remain indistinguishable.

Case $\alpha_M \epsilon < \gamma_M B < \alpha_M(2P - \epsilon)$ In the case with $\alpha_M \epsilon < \gamma_M B < \alpha_M(2P - \epsilon)$ and no effort, the S-manager and (thus) employees of the S-firm choose X_h unless both $\mathcal{B}_{X_h} = 0$ and $\mathcal{B}_{X_l} = B$, in which case they choose X_l. It follows that with probability $1/2$, the manager chooses X_h with $\mathcal{B}_Z = B$;

with probability 1/6, she chooses X_h with $\mathcal{B}_Z = 0$ and with probability $2/6 = 1/3$, she chooses X_l with $\mathcal{B}_Z = B$. (For these probabilities, note the following. The likelihood that a permutation of $\{B, B, 0, 0\}$ has a $+B$ in the first position is obviously 1/2. Conditional on having 0 in the first position, however, the second position equals 0 with probability 1/3 and $+B$ with probability 2/3.) So, conditional on execution, the expected monetary profit is $P - \epsilon/3$, while the expected social impact equals $5B/6$. Moreover, with probability 1/3, the firm distinguishes itself as an S-firm. It follows now that an A-employee gets payoff $w+\underline{\psi}\alpha_E P$ from an A-firm and $w+\underline{\psi}[\alpha_E(P-\epsilon/3)+[qU(\mu,\nu)+(1-q)U(0,\nu)]/3+2U(0,0)/3)]$ from an S-firm. An S-employee gets payoff $w+\underline{\psi}\alpha_E P+(1-q)U(1,0)$ from an A-firm and $w+\underline{\psi}[\alpha_E(P-\epsilon/3)+5\rho\gamma_E B/6+[qU(\mu,\nu)+(1-q)U(1,\nu)]/3+2[qU(0,0)+(1-q)U(1,0)]/3]+(1-\underline{\psi})(1-q)U(1,0)$ from an S-firm. In any core allocation, all employees in an S-firm are S-employees whereas employees of A-firms are randomly drawn. It follows that S-employees in the S-firm get net utility (relative to the outside option) of $w+\underline{\psi}[\alpha_E(P-\epsilon/3)+5\rho\gamma_E B/6+(U(1,1)-(1-q)U(1,0))/3]$ so that the wage equals $w = -\underline{\psi}[\alpha_E(P-\epsilon/3)+5\rho\gamma_E B/6+[U(1,1)-(1-q)U(1,0)]/3]$. The profit of the S-firm thus equals $\underline{\psi}[(P-\epsilon/3)+K5\rho\gamma_E B/6+K[U(1,1)-(1-q)U(1,0)]/3]$. The expected profits of the A-firms are again $\underline{\psi}P$. The S-firm is thus indeed more profitable than A-firms at low ϵ, high $\gamma_E B$, or high $U(1,1)$.

The case with effort is completely similar because for most values effort increases existing differences in payoffs (and in no case changes the qualitative nature of the equilibrium). Both A- and S-employees in A-firms will exert the same effort. S-employees in the S-firm will exert more effort than (hypothetical) A-employees in the S-firm. The core allocation of employees thus remains the same, with only S-employees in the S-firm. And these S-employees in the S-firm will exert more effort than employees in the A-firm if and only if the employees' net expected utility was higher, thus indeed typically increasing existing differences.

Consider now the special case (for $\alpha_M\epsilon < \gamma_M B < \alpha_M(2P-\epsilon)$) with $U = \mu+\nu$ (i.e., $\lambda = 1$), $\rho\gamma_E B = 0$, and no effort. An A-employee gets payoff $w+\underline{\psi}\alpha_E P$ from an A-firm and $w + \underline{\psi}[\alpha_E(P - \epsilon/3) + [q(\mu + \nu) + (1 - q)(0 + \nu)]/3]$ or $w + \underline{\psi}[\alpha_E(P - \epsilon/3) + [q\mu + \nu]/3]$ from an S-firm. An S-employee gets payoff $w + \underline{\psi}\alpha_E P + (1 - q)(1 + 0)$ from an A-firm and $w + \underline{\psi}[\alpha_E(P - \epsilon/3) + [q(\mu + \nu) + (1 - q)(1 + \nu)]/3 + 2[(1 - q)(1 + 0)]/3] + (1 - \underline{\psi})(1 - q)(1 + 0)$ or $w + \underline{\psi}[\alpha_E(P - \epsilon/3) + [q\mu + \nu + (1 - q)]/3 + 2(1 - q)/3] + (1 - \underline{\psi})(1 - q)$ or

$w + \underline{\psi}[\alpha_E(P - \epsilon/3) + [q\mu + \nu]/3)] + (1 - q)$ from an S-firm. It follows that the S-employee always gets $1 - q$ more than the A-employee, independent of which firm (or which other employees) the S-employee is allocated to, so that allocating S-employees to S-firms is now no more efficient than allocating them to A-firms. It further follows that any allocation can be in the core and that there is no sorting in equilibrium. But that further implies that $\mu = \nu = 0$ and thus the payoffs for the S-firm become lower for both types of employees than for the A-firm. This finally implies that the S-firm is less profitable than the A-firms. And, as before, considering also effort preserves these results as the net utility of A- and S-types is the same for the same firm, so that they exert the same effort in the same firm (but more effort in A-firms than in S-firms, accentuating the profit difference).

Case $\gamma_M B > \alpha_M(2P - \epsilon)$ When $\gamma_M B > \alpha_M(2P - \epsilon)$, an S-manager chooses the highest paying action with social impact $+B$ (and thus private benefit $+\gamma_M B$). Some accounting implies that the manager will choose X_h with probability 1/2, X_l with probability 1/3, and Y_l with probability 1/6, giving an expected payoff of $2P/3 - \epsilon/6$ and an expected social impact of B. The S-firm distinguishes itself from A-firms with probability 1/2.

In this case and without effort, an A-employee gets payoff $w + \underline{\psi}\alpha_E P$ from an A-firm and $w + \underline{\psi}[\alpha_E(2P/3 - \epsilon/6) + [qU(\mu, \nu) + (1 - q)U(0, \nu)]/2]$ from an S-firm. An S-employee gets payoff $w + \underline{\psi}\alpha_E P + (1 - q)U(1, 0)$ from an A-firm and $w + \underline{\psi}[\alpha_E(2P/3 - \epsilon/6) + \rho\gamma_E B + [qU(\mu, \nu) + (1 - q)U(1, \nu)]/2 + [qU(0, 0) + (1 - q)U(1, 0)]/2] + (1 - \underline{\psi})(1 - q)U(1, 0)$ from an S-firm. In any core allocation, all employees in an S-firm are S-employees. It follows that S-employees in an S-firm get net utility (relative to the outside option) of $w + \underline{\psi}[\alpha_E(2P/3 - \epsilon/6) + \rho\gamma_E B + [U(1, 1) - (1 - q)U(1, 0)]/2]$ so that the wage equals $w = -\underline{\psi}[\alpha_E(2P/3 - \epsilon/6) + \rho\gamma_E B + [U(1, 1) - (1 - q)U(1, 0)]/2]$. The profit of the S-firm thus equals $\underline{\psi}[(2P/3 - \epsilon/6) + K\rho\gamma_E B + K[U(1, 1) - (1 - q)U(1, 0)]/2]$. The profit of the A-firm is, as always, $\underline{\psi}P$, which may be higher or lower than the S-profit. Relative to the earlier case with $\alpha_M\epsilon < \gamma_M B < \alpha_M(2P - \epsilon)$, the S-profit is now indeed lower given the assumption that $P - \epsilon/2 > \max\{\gamma_M B, \gamma_E B, KU(1, 1)\}$, showing that the highest S-profit is at intermediate $\gamma_M B$. Including effort preserves the qualitative characteristics of the equilibrium, though it will not necessarily be the case any more that profits of the S firm are lower in this case than in the case with $\alpha_M\epsilon < \gamma_M B < \alpha_M(2P - \epsilon)$ (as the effort of the current case may be higher, for example when $\alpha_E \downarrow 0$).

Consider now again the case (for $\gamma_M B > \alpha_M(2P-\epsilon)$) with $U = \mu+\nu$, $\rho\gamma_E B = 0$, and no effort. An A-employee gets payoff $w + \underline{\psi}\alpha_E P$ from an A-firm and $w+\underline{\psi}[\alpha_E(2P/3-\epsilon/6)+[q(\mu+\nu)+(1-q)\nu]/2]$ or $w+\underline{\psi}(\alpha_E(2P/3-\epsilon/6)+[q\mu+\nu]/2)$ from an S-firm. An S-employee gets payoff $w + \underline{\psi}\alpha_E P + (1-q)(1+0)$ from an A-firm and $w+\underline{\psi}[\alpha_E(2P/3-\epsilon/6)+[q(\mu+\nu)+(1-q)(1+\nu)]/2+[(1-q)(1+0)]/2]+(1-\underline{\psi})(1-q)(1+0)$ or $w+\underline{\psi}[\alpha_E(2P/3-\epsilon/6)+[q\mu+\nu+(1-q)]/2+(1-q)/2]+(1-\underline{\psi})(1-q)$ or $w+\underline{\psi}[\alpha_E(2P/3-\epsilon/6)+[q\mu+\nu]/2]+(1-q)$ from an S-firm. It follows that the S-employee always gets $1-q$ more than the A-employee, independent of how the S employee is allocated, so that allocating S-employees to S-firms is now no more efficient than allocating them to A-firms. It further follows that any allocation can be in the core and that there is no sorting in equilibrium. But that further implies that $\mu = \nu = 0$ and thus the payoffs for the S-firm become lower for both types of employees than for the A-firm. This finally implies that the S-firm is less profitable than the A-firms. And, per an analogous argument to before, including effort preserves these results.

■

4 Model where managers care about reputation

To see how the results may extend to a setting where managers also care about reputation and identity, consider the earlier model but with the following assumptions and modifications:

- Managers get (apart from the wage $w_M = 0$ and profit share $\alpha_M\Pi$) utility $U(\mu,\nu)$ from identity and reputation. Like employees, managers also forget with probability q everything that is not public or common knowledge. We also assume that, for all players, $U(\mu,\nu) = \mu\nu$.

- There are only two actions, X and Y, with respective monetary payoffs P and 0 and respective social impact 0 and B. (Note that this thus assumes perfect negative correlation between $\mathcal{P}_Z$ and $\mathcal{B}_Z$.) The social impact B gives private benefits $\gamma_E B$ and $\gamma_M B$, as before.

- Consider the case with no effort and $\underline{\psi} = 1$, and assume $q > \alpha_M P$ (so that it is worthwhile for an A-manager to imitate an S-firm if that would give her $\mu = \nu = 1$ upon forgetting her type).

It is fairly easy to show that it is again possible for the S-firm to be more profitable overall than the A-firms. To see this, note the following. It is straightforward to show that for any firm that chooses X, the beliefs will be $\mu = \nu = 0$. Let now the belief for a firm that chooses Y be $\hat{\nu} \in [0,1]$. (Note that this will be both μ and ν when an employee forgets her own type.) Once wages are sunk, an A-manager then expects personal payoff $\alpha_M P$ from choosing X and $q\hat{\nu}^2$ from choosing Y. An S-manager gets $\alpha_M P$ from choosing X and $\gamma_M B + q\hat{\nu}^2 + (1-q)\hat{\nu}$ from choosing Y. It is fairly easy to see that there is no symmetric equilibrium in pure strategies (as A-managers will want to imitate S-managers): every symmetric equilibrium must have mixed strategies for the A-firms, with the S-firm always choosing Y. The equilibrium must thus be such that an A-manager is indifferent, i.e., $\alpha_M P = q\hat{\nu}^2$ or $\hat{\nu} = \sqrt{\alpha_M P/q}$. Let τ be the probability that an A-firm chooses Y, then in equilibrium it must be that $\tau = \frac{1-\hat{\nu}}{\hat{\nu}(N-1)}$. Note that this probability converges to zero in the limit as $N \to \infty$. In terms of utilities, an A-employee in an A-firm has expected utility $w+(1-\tau)\alpha_E P+\tau q\hat{\nu}^2$ which equals $w+\alpha_E P$ in the limit. An A-employee in an S-firm has utility $w+q\hat{\nu}^2 = w+\alpha_M P$. An S-employee in an A-firm has expected utility $w+(1-\tau)\alpha_E P+\tau(q\hat{\nu}^2+(1-q)\hat{\nu})$, which equals again $w+\alpha_E P$ in the limit (as $\tau \downarrow 0$ when $N \to \infty$). An S-employee in an S-firm, finally, has expected utility $w+\rho\gamma_E B+q\hat{\nu}^2+(1-q)\hat{\nu}$ or $w+\rho\gamma_E B+\alpha_M P+(1-q)\sqrt{\alpha_M P/q}$. So whereas an S employee has the same expected utility as an A-employee from working in an A firm, she has relatively higher expected utility from working in an S-firm. It follows that it is efficient for employees to sort. In any core solution, all employees of the S-firm will be S-employees with wage $w = -(\rho\gamma_E B+\alpha_M P+(1-q)\sqrt{\alpha_M P/q})$. An A -firm's profit (prior to the manager's pay) is simply P whereas that of an S-firm equals $K(\rho\gamma_E B+\alpha_M P+(1-q)\sqrt{\alpha_M P/q})$. It follows that the S-firm may (but not necessarily will) outperform the A-firms, depending on the parameters.