Open access • Journal Article • DOI:10.2139/SSRN.645741

Moral hazard and moral motivation: Corporate social responsibility as labor market screening — Source link

Kjell Arne Brekke, Karine Nyborg, Karine Nyborg

Institutions: University of Oslo, Institute for the Study of Labor

Published on: 01 Oct 2004 - Social Science Research Network (Oslo: University of Oslo, Department of Economics)

Topics: Moral disengagement, Social responsibility, Corporate social responsibility and Moral hazard

MEMORANDUM

No 25/2004

Moral hazard and moral motivation: Corporate social responsibility as labor market screening

Kjell Arne Brekke and Karine Nyborg

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the **University of Oslo** **Department of Economics**	In co-operation with **The Frisch Centre for Economic Research**
P. O.Box 1095 Blindern N-0317 OSLO Norway Telephone: + 47 22855127 Fax: + 47 22855035 Internet: http://www.oekonomi.uio.no/ e-mail: econdep@econ.uio.no	Gaustadalleén 21 N-0371 OSLO Norway Telephone: +47 22 95 88 20 Fax: +47 22 95 88 25 Internet: http://www.frisch.uio.no/ e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 24	Alexander W. Cappelen and Bertil Tungodden Local autonomy and interregional equality. 24 pp.
No 23	Jo Thori Lind Does permanent income determine the vote?. 32 pp.
No 22	Erik Biørn, Terje Skjerpen and Knut R. Wangen Can Random Coefficient Cobb-Douglas Production Functions Be Aggregated to Similar Macro Functions?. 31 pp.
No 21	Atle Seierstad Open mapping theorems for directionally differentiable functions. 13 pp.
No 20	Contract Renewal Helge Holden, Lars Holden, and Steinar Holden. 44 pp.
No 19	Jo Thori Lind Repeated surveys and the Kalman filter. 19 pp.
No 18	Kari Eika When quality today affects service needs Tomorrow. 30 pp.
No 17	Rolf Golombek and Michael Hoel Unilateral emission reductions when there are cross -country technology spillovers. 23 pp.
No 16	Kjetil Bjorvatn and Alexander W. Cappelen Globalisation, inequality and redistribution. 17 pp.
No 15	Alexander W. Cappelen and Bertil Tungodden Rewarding effort. 31 pp.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

Moral hazard and moral motivation:

Corporate social responsibility as labor market screening

Kjell Arne Brekke and Karine Nyborg

October 1, 2004

Abstract

Morally motivated individuals behave more cooperatively than predicted by standard theory. Hence, if a firm can attract workers who are strongly motivated by ethical concerns, moral hazard problems like shirking can be reduced. We show that employers may be able to use the firm's corporate social responsibility profile as a screening device to attract more productive workers. Both pooling and separating equilibria are possible. Even when a substantial share of the workers have no moral motivation whatsoever, such screening may in fact drive every firm with a low social responsibility profile out of business.

Keywords: Self-image, teamwork, shirking, voluntary abatement.

JEL codes: D21, D62, D64, J31, Q50, Z13.

Address: The Ragnar Frisch Centre for Economic Research, Gaustadalléen 21, N-0349 Oslo, Norway (both authors). E-mail addresses: karine.nyborg@frisch.uio.no (Nyborg, corresponding author), k.a.brekke@frisch.uio.no (Brekke).

Acknowledgements: This paper is part of the project "Green consumers and producers", a collaboration between the Ragnar Frisch Centre for Economic Research (project # 3113) and Center for Development and the Environment at University of Oslo. Funding from the Research Council of Norway through the *SAMSTEMT* program is gratefully acknowledged.

1 Introduction

Many private firms make a considerable effort to be, or at least to appear, socially responsible. They contribute to charity, invest in costly abatement equipment even when pollution would have been legal, or commit themselves voluntarily to ethical principles increasing their production costs, such as abstaining from the use of child labor in developing countries.[1] But why would a private firm pay to promote social values? And if a firm incurs extra costs for the sake of social responsibility, will it not be wiped out of the market by less responsible competitors?

In the present paper, we will analyze conditions under which some firms voluntarily, without public intervention, choose to undertake costly measures promoting social goals, for example pollution prevention. It has been pointed out previously that such behavior may be profitable for firms if customers have an extra willingness to pay for products produced in a "responsible" way (Arora and Gangopadhyay 1995, Moon et al. 2002, Björner et al. 2004); or if firms expect that they can pre-empt the introduction of taxes or regulations (Maxwell et al. 2000). While we acknowledge the relevance of these explanations, we will disregard both of them in the analysis below, in order to focus on another, but less studied, reason for firms to pursue social goals, namely that of worker motivation.

Recent research has demonstrated that a substantial number of individuals behave more cooperatively than predicted by standard economic models (see, e.g., Fehr and Falk 2002, Schram 2000; for theoretical analyses see e.g. Sudgen 1984, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, or, for a survey, Nyborg and Rege 2003). Such individuals are attractive partners in economic interaction: They cheat less, they exert more effort, and they contribute more to public goods. The problem is, of course, that it is hard to observe a priori whether an individual is a "cooperative type" or not.[2]

[1] For example, in the spring of 2004, Exxon, Chiquita, McDonald's, Coca-Cola and Ford Motor Company all had statements of committment to environmental and social values figuring prominently on their homepages, together with reports of costly measures taken to promote these values.

[2] Frank (1988) suggests that this has been so important throughout history that humans have evolved observable biological signals of intention, like blushing when telling a lie.

Below, we will demonstrate that firms may be able to use their social responsibility profile to attract more reliable workers. In long-term equilibrium, firms with high and low social responsibility may co-exist. More strikingly, firms with low social responsibility could be driven entirely out of business, even when a share of the workers demand no wage compensation at all to work in such firms.

The crux of our analysis will be an assumption that cooperative behavior in various situations originates from a common underlying principle of ethics. We also assume that the strength of moral motivation differs between individuals. Then, the behavior of any given individual in one context provides an indication of the behavior to expect from her in another context. Firms can exploit this correlation to attract workers who shirk less than others, by combining a high level of social corporate responsibility (for example, environment-friendly production) with relatively low wages. The logic of our argument has much in common with that of standard signalling and/or screening models (Spence 1973, Stiglitz 1975): the only workers who will find the above deal attractive are those who have a relatively high level of moral motivation.

The long-term market equilibrium in our model can either be characterized by pooling (only brown firms, or only green firms) or separation (some green and some brown firms, and the workers with strongest moral motivation being employed by the green firms). This will depend on the costs of being socially responsible, and on the strength and heterogeneity of workers' moral motivation.

The empirical findings reported in Frank (2003) indicate that many individuals do in fact prefer their employer to be socially responsible and, moreover, may be prepared to pay a substantial premium to achieve this goal. Using data for Cornell graduates, Frank found that even after controlling for sex, curriculum, and academic performance, employees of for-profit firms in his sample earned 59 percent more, on average, than employees of non-profit firms. Rating the social responsibility image of each employer through an interview survey, he found a large and statistically significant compensating salary differential, with the jobs rated as *less* socially responsible earning substantially *higher* wages. Frank also asked respondents to choose between pairs of hypothetical jobs, where the nature of the work was similar while the employers' social responsibility reputation was different. After picking their preferred job from each pair, subjects were asked to state the wage differential required to make them reverse their choice. The results were striking: For example, 88 percent preferred to work as an ad copywriter for the American Cancer Society rather than for Camel Cigarettes, and the average reported switching premium was, in this case, as high as $24,333 per year.[3]

While these numbers are substantial, the main mechanism of our model requires only that when choosing their employer, a positive share of workers prefer firms with high social responsibility to those with low social responsibility, *ceteris paribus*. The wage compensation required to make workers prefer the low responsibility firm need not necessarily be large.

In what follows, our intuitive understanding of "moral motivation" is taken from Brekke, Nyborg and Kverndokk (2003), although our formalization will differ on certain points. Individuals are assumed to have preferences for a good self-image, in addition to other private and public goods. Further, they are assumed to assess their self-image through considering the consequences for social welfare if everybody acted like themselves; a simplified version of Kant's categorical imperative. In contrast to Brekke et al. (2003), however, we will allow for heterogeneity in individuals' moral motivation, and explore the consequences of this in long-term market equilibrium.

Corporate social responsibility has been defined by the EU Commission as “a concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with their stakeholders on a voluntary basis” (EU Commission 2002, p.5). In the formal analysis below, the firm's choice of social responsibility profile is assumed to be discrete: Either it pays a fixed social responsibility cost, and is termed "green", or it does not pay this cost and is termed "brown". We will begin by presenting our formalization of moral motivation, and its implication for worker behavior, and then proceed to study market equilibrium and the requirements for arriving at pooling or separating equilibria.

[3] The median premium was lower, but still a substantial $15,000 per year.

2 Workers' moral motivation

Consider an economy characterized by a perfectly competitive labor market and full employment. Assume that there are N workers with identical utility functions

$$U_i = u(x_i, E, e_i) + S_i \tag{1}$$

where x_i is individual i's private consumption, measured in monetary units, E is a pure public good, which we can think of as environmental quality, and e_i is a measure of the effort the individual exerts while at work. The utility function is increasing in consumption and environmental quality, and decreasing in effort. S_i is a measure of benefits derived from keeping a self-image as a socially responsible individual. Additive separability in the latter variable is assumed for the case of simplicity. Further, u is assumed to be quasiconcave. Workers choose in which firm to seek employment, and then, given their employer, how much effort to exert while at work.

Firms are characterized by team production with unobservable (or at least unverifiable) individual effort. Individual wages are equal for all workers within a given firm, since employers cannot distinguish between different workers' effort.[4] Any worker i's income is given by the wage per worker offered from i's employer. Below, we will assume that every firm uses the same technology, but that firms may differ with respect to corporate social responsibility profile, such that each firm can be denoted either "green" or "brown". In equilibrium, the offered wage may differ between green and brown employers. Hence the individual's budget constraint is given by

$$x_i = w(\tau_i). \tag{2}$$

where τ_i is the corporate social responsibility profile of the firm where i chooses to work, and $w(\tau_i)$ is i's

[4] Holmstrom (1982) shows that moral hazard problems in teams could, in principle, be solved through incentive schemes involving group penalties. Below, we will assume that workers regard their own contribution to average productivity as negligible. This implies that, in contrast to Holmstrom's model, group penalties will not be effective.

wage.[5] Let $\tau_i = 0$ if i works for a brown firm, and $\tau_i = 1$ if i works for a green firm.

Workers have their own views of what a good society is like. To evaluate social welfare, they use their own subjective perception of a social welfare function (see Brekke, Lurås and Nyborg 1996, Nyborg 2000). To make everything as simple as possible, let us start by assuming that every worker has a utilitarian social welfare function of the following type:

$$V = \sum_{j=1}^{n} U_j \tag{3}$$

Individuals evaluate their self-image as socially responsible individuals, S_i, by asking themselves: "What would happen to social welfare if everybody acted like me?" The more favorable the answer to this question is – i.e., the better the social welfare consequences if everybody made the same choices as i – the better is i's self-image. Such a principle for judging one's own moral stance is consistent with well-known and widely accepted ethical views, such as the Biblical assertion that you should treat others as you would want others to treat yourself (Mathew 7.12), or Immanuel Kant's categorical imperative (to act only according to those maxims that can be consistently willed as a universal law (Audi 1995, p. 403)).[6]

In the present model, a worker has two choices to make: Which type of firm to work for (τ_i), and how much effort to exert (e_i). Thus, let $\widetilde{V}(e_i, \tau_i)$ denote social welfare (3) *if everybody made the same choices as* i. Further, let $\alpha_i \in [0, \bar{\alpha}]$, where $\bar{\alpha} < 1$, be an individual-specific parameter indicating the importance of

[5] To keep the analysis simple, we will discuss this as if each individual works full-time in one and only one firm. However, in the formal analysis below, we must allow the marginal worker in each firm to share his time between two employers; hence there may be *one* worker who is partly employed by a green firm and partly by a brown firm. For the sake of simplicity, we will ignore this complication below. Since the economy is large, and workers are assumed to disregard the effect of their own effort on equilibrium wages, this simplification does not substantially affect our results.

[6] In a Norwegian survey from 1999, 88 percent of those who recycled household waste agreed or agreed partly to the following statement: "I recycle partly because I think I should do what I want others to do". See Brekke et al. (2003) and Bruvoll et al. (2002).

social welfare considerations for individual i's self-image. Then self-image is determined as[7]

$$S_i = \alpha_i(\widetilde{V}(e_i, \tau_i)) \tag{4}$$

Note that others' self-image will affect this hypothetical socal welfare measure, through those others' utility functions. The reader might suspect that this produces a kind of circular reasoning, implying that utility and welfare may not be well-defined. However, Lemma 1 establishes that individual utility can indeed be defined without reference to others' self-image. The intuitive reason is that if everybody else acted just like i, those others' self-image would be affected in similar ways as i's too, the only difference being the proportionality factor α_i.

Assume that $\sum_{j=1}^{n} \alpha_j < 1$. Let $\widetilde{Y}(e_i, \tau_i)$ denote the value of any variable Y in the hypothetical case that $e_i = e_j$ and $\tau_j = \tau_i$ for all $i, j \in \{1, ..., N\}$. Then we have the following:

Lemma 1 *Individual utility can be written as*

$$U_i = u(x_i, E, e_i) + \alpha_i N K u(\tilde{x}_i(e_i, \tau_i), \tilde{E}(e_i, \tau_i), e_i) \tag{5}$$

where $K = 1/[1 - \sum_{j=1}^{n} \alpha_j] > 0$.

Proof. See Appendix. ■

Hence, with these assumptions, it is not essential whether self-image benefits are taken into account in the social welfare evaluation or not. In the economic literature, there has been quite some discussion of whether "altruistic benefits" should be counted in social welfare calculations (see, for example, Milgrom 1993, Johansson 1992). With the specification used here, individuals' benefit from "being moral" at least does not affect their assessments of what is in fact morally right.

[7] This formalization is different from that of Brekke et al. (2003). There, the question "what would happen to social welfare if everybody acted like me?" was used to identify the morally ideal *contribution*, while self-image was determined by the distance between this ideal contribution and one's actual contribution. Here social welfare calculations enter more directly. While the underlying ethical principle is very similar, the latter approach is simpler and better suited for analysis of *multiple* social dilemmas.

3 Firms and pollution

Suppose that the cost-minimizing production technology is well-known and available to everyone. Then, entry and exit from the industry will ensure that in long-term equilibrium, there will be no rents. The production function, $y^h = f(L^h)$, where y^h is production in firm h, is thus assumed to be the same for all firms, with $f' > 0$ and $f'' > 0$. L^h is effective labor input in firm h, which depends on the number of employees in firm h, N^h, and their effort, in the following way:

$$L^h = N^h(1 + \mu \overline{e}^h) \tag{6}$$

where $\overline{e}^h = (\sum_i e_i)/N^h$ and the sum is over all i who are employed by firm h (i.e. average effort level among the firm's workers), and $\mu > 0$ is a parameter that determines the impact of effort on effective labour. If a worker i is just present at work, displaying the minimum performance consistent with not being fired, then $e_i = 0$.[8] If a worker exerts more effort than this minimum level, $e_i > 0$. Hence e_i can be regarded as the worker's voluntary and "unpaid" contribution to the firms' productivity, as workers' salary is fixed. Since individual effort is unobservable, the firm cannot differentiate wages according to individual effort, and consequently faces a moral hazard problem.

Assume that an end-of-pipe abatement technology is available at a fixed cost A, reducing firms' polluting emissions to zero (i.e. a level where environmental damages do not occur) if installed. Hence, there can be two types of firms: green firms pay A and do not pollute, while brown firms use the same production technology, but have not installed the cleaning equipment, so they pollute the environment, and have lower fixed costs. Below, let τ^h reflect whether firm h has chosen to be green, so that $\tau^h = 1$ if firm h is green, and $\tau^h = 0$ if firm h is brown.[9]

In long-term equilibrium, each firm type (if existing) has zero profit, so revenues must equal costs. In addition to labor costs and abatement costs, we assume that there is a fixed cost F in production, reflecting

[8] This implies that there must be some minimum effort level below which even lower effort would be observable.

[9] Note that while τ^h denotes the *firm's* choice of "greenness", τ_i denotes the *worker's* choice. If $\tau_i = 0$, then we also have $\tau^h = 0$ if i works for firm h.

capital costs. Normalizing the product price to 1, this implies

$$y^h = N^h W^h + F + \tau^h A \tag{7}$$

for every firm h, where W^h is the wage per worker offered by the firm. This implies that in equilibrium, this wage must equal total production value per worker, after subtraction of the fixed costs:

$$W^h = [f(N^h(1 + \mu \bar{e}^h)) - F - \tau^h A]/N^h \tag{8}$$

Since green firms must cover their abatement costs, such firms may offer a lower wage. We assume that firms maximize their profits taking equilibrium wages for each firm type as exogenously given, but acknowledging that the equilibrium wage may be different for green and brown firms:

$$W^h = w(\tau^h) \tag{9}$$

For any τ^h, profit maximization with respect to the number of workers yields the first-order condition for profit maximization, requiring that the productivity of the marginal worker equals his wage:

$$f'(N^h(1 + \mu \bar{e}^h))(1 + \mu \bar{e}^h) = w(\tau^h) \tag{10}$$

Equations (8) and (10) imply that in equilibrium, revenues less fixed costs per worker equals the value of marginal productivity. This condition will determine the number of firms in equilibrium.

The optimal labor stock in each firm, N^h, depends on workers' average effort, but in equilibrium, *effective* labor input (L^h) in each firm does not, as established by the following Lemma. This implies that in each firm type, wages are increasing in workers' average effort:

Lemma 2 *Assuming zero profit in long run equilibrium, it follows that wages are set at*

$$w(\tau^h) = f'(L(\tau^h))(1 + \mu \bar{e}^h) \tag{11}$$

where $L(\tau^h)$*,the effective labor input for a firm of type* τ^h*, is fixed, independent of workers' efforts.*

Proof. See Appendix. ■

If workers' effort increase, new firms will be established, so total production is affected by effort; but production *per firm* is not.[10]

For green firms there is no pollution, while for brown firms there is a fixed emission coefficient $z > 0$ per unit of production. Environmental quality is deteriorated by pollution from firms in the following way:

$$E = E^0 - \sum_{h=1}^{H}([1 - \tau^h]zf(L^h)) \tag{12}$$

where H is the total number of firms. If there is no pollution, E will stay at a base level E^0.

4 What would happen if everybody acted like me?

The worker uses his knowledge about society and his perception of the social welfare function to determine the morally relevant aspects of his choice. As discussed above, the worker has two choices to make: Whether to work for a green or a brown firm, and what effort level to exert. When making his choice of firm type, the worker compares the maximum utility level he can achieve in green and brown firms, respectively. Since utility depends on self-image, which in turn depends on the welfare consequences if everybody acted like i, $\widetilde{V}(e_i, \tau_i)$, we will first explore the properties of the latter function.

We assume that each worker i considers average effort, wage levels, and environmental quality as exogenous; i.e., workers do not believe that their own behavior has a perceptible effect on these variables. However, if everybody behaved just like i, i.e. $e_i = e_j$ and $\tau_j = \tau_i$ for all $i, j \in \{1, ..., N\}$, the effects could of course not be neglected.

If everybody increased their effort marginally, this would increase the equilibrium wage, and thus consumption, by f'. It would also increase everybody's disutility of effort. If the worker is employed by a brown firm, however, the increased total production would be accompanied by a reduced environmental quality. The following Lemma establishes the effects on $\widetilde{V}(e_i, \tau_i)$ of a marginal increase in e_i.

[10] Note that $L(\tau^h)$ is effective labor input *per firm* of type τ^h. *Total* effective labor input in green and brown firms, respectively, depends on the total number of firms, which varies endogenously.

Lemma 3 *When e_i increases marginally, the resulting change in $\widetilde{V}(e_i,\tau_i)$ depends on the social benefits of changed consumption, environmental quality and effort levels if everybody behaved like i:*

$$\partial\widetilde{V}(e_i,\tau_i)/\partial e_i = NK(\mu u'_x f' - \mu u'_E([1-\tau_i]zf(L(0))\frac{N}{L(0)}) + u'_e) \tag{13}$$

where $L(0)$ is effective labor input per brown firm.

Proof. See Appendix. ■

Before proceeding, it is useful to make sure that "green" always corresponds to "socially responsible" within the current model. If, for example, cleaning equipment were so expensive that workers judged abatement to be socially inferior to no abatement, they would consider *brown* technology as the most socially responsible choice. Hence, throughout our analysis the following two conditions will be assumed to hold:

$$\widetilde{V}(e_i,1) > \widetilde{V}(e_i,0) \tag{14}$$

$$\partial\widetilde{V}(e_i,1)/\partial e_i > \partial\widetilde{V}(e_i,0)/\partial e_i. \tag{15}$$

5 Workers' utility maximization

Recall that when maximizing utility, a worker considers average effort in any given firm to be exogenous, and the same holds for wage rates in each firm type and for environmental quality. It is only in the ethical consideration reflected in $\widetilde{V}(e_i,\tau_i)$ that it is relevant for him to regard these variables as endogenous (in the special sense discussed above).

Differentiating (1) with respect to e_i, taking τ_i as given, yields the following first order condition for utility maximization:

$$\alpha_i\partial\widetilde{V}(e_i,\tau_i)/\partial e_i = -u_e$$

Inserting from (13) gives

$$\alpha_i NK(\mu u'_x f' - \mu u'_E([1-\tau_i]zf(L(0))\frac{N}{L(0)}) + u'_e) = -u_e. \tag{16}$$

Assuming an interior solution, the optimal effort level for a morally motivated worker is the level where marginal benefits of effort, in terms of a better self-image, just equals the marginal disutility of effort. Note that the worker thus provides less effort than that he would consider morally best: To maximize the hypothetical social welfare if everybody acted like him, he should choose e_i such that $\partial\widetilde{V}(e_i^*, \tau_i)/\partial e_i = 0$, while utility maximization implies $\partial\widetilde{V}(e_i, \tau_i)/\partial e_i = -u_e/\alpha_i > 0$. Hence, although the worker stretches himself towards his conception of a morally ideal behavior, he stops short of reaching that ideal.

As long as $\partial\widetilde{V}(e_i, \tau_i)/\partial e_i > 0$, individuals with a stronger moral motivation – that is, individuals with a higher α_i – receive a higher marginal compensation for their effort than others, in terms of improved self-image. Nevertheless, without further restrictions on the utility function we cannot guarantee that effort is increasing in α_i, since there will in general be both income effects and substitution effects. With reasonable assumptions, however, workers with a stronger moral motivation will indeed exert more effort. This holds, for example, if the utility function is linearly separable and strictly concave in e_i. The proposition below specifies conditions under which effort is increasing in the moral motivation parameter α_i.

Proposition 1 *Assume that $u(x_i, E, e_i) = \widehat{u}(x_i, E) + \overline{u}(e_i)$, where $\widehat{u}'_x > 0, \widehat{u}'_E > 0, \overline{u}'(0) = 0$ and $\overline{u}'' < 0$. Then (a) for $\tau_i = 1$, e_i is strictly increasing in α_i, (b) for $\tau_i = 0$, then either $\partial\widetilde{V}(0,0)/\partial e_i \leq 0$ and $e_i = 0$ for all α_i or $\partial\widetilde{V}(0,0)/\partial e_i > 0$ and e_i is strictly increasing in α_i. Finally (c) any given worker with $\alpha_i > 0$ will exert more effort if he works in a green firm than if he works in a brown firm.*

Proof. See Appendix. ■

When workers are morally motivated, and $\partial\widetilde{V}(0,0)/\partial e_i > 0$, then workers choose $e_i > 0$. Hence, not unexpectedly, moral motivation alleviates the moral hazard problems in team production pointed out by Holmstrom (1982). However, in brown firms this effect is partly offset by the worker's concern that he contributes to pollution: Any strictly positive effort level yields lower self-image benefits than it would had he worked for a green firm.

Consequently, morally motivated workers will have a strictly positive willingness to pay to work in a green rather than a brown firm; so in equilibrium, green firms may be able to hire workers at a lower wage than brown firms. This willingness to pay, let us denote it ϕ_i, can be defined implicitly as

$$u(w(0) - \phi_i, E, e_i^1) + \alpha_i(\widetilde{V}(e_i^1, 1)) = u(w(0), E, e_i^0) + \alpha_i(\widetilde{V}(e_i^0, 0)) \tag{17}$$

where e_i^τ should be interpreted as the effort individual i would provide if working in a firm of type $\tau \in \{0, 1\}$ (recall that 0 denotes brown and 1 green).

Lemma 4 *With the assumptions specified in Proposition 1, ϕ_i is a strictly increasing function of α_i.*

Proof. See Appendix. ■

It may seem trivial to point out that a substantial willingness to pay for working in green firms can enable such firms to survive in the long run, in spite of abatement costs, due to lower wages. Nevertheless, as we will demonstrate below, workers' willingness to pay need not be substantial for green firms to survive. The crucial feature is only that *if pay were equal in both firm types, some fraction of the workers would strictly prefer green firms.* Even with a quite marginal level of willingness to pay, this allows for labor market screening; consequently, green firms can survive, not just due to lower wages, but because their workers are more productive. Let us now turn to this issue.

6 Attracting workers with a high work morale

All else equal, a profit maximizing firm prefers to hire workers with a high moral motivation, since these workers exert higher effort (or, equivalently, shirk less) when productivity cannot be observed. The problem is, of course, how can the firm know who is morally motivated?

Let $\phi(\alpha_i)$ be i's willingness to pay for working in a green rather than a brown firm, as a function of the worker's moral motivation parameter α_i. A worker i prefers working in a green firm if[11]

$$w(1) \geq w(0) - \phi(\alpha_i). \tag{18}$$

If green firms pay a strictly lower wage than brown firms, the only applicants to jobs in green firms will be those who have a sufficiently high $\phi(\alpha_i)$. Since not only willingness to pay, but also the worker's effort

[11] Recall that $w(1)$ is the wage in green firms and $w(0)$ the wage in brown firms.

level, increase in α_i, firms may be able to use a green social responsibility profile, combined with a relatively low wage, as a screening device to attract more productive workers.

To see this, let α denote a threshold such that all i with $\alpha_i \geq \alpha$ prefer to work in a green firm, and all i with $\alpha_i < \alpha$ prefer a brown firm.[12] For both firm types, average productivity varies with the level of α: A very high α, for example, means that there are only a few green firms, but since they employ workers with unusually strong moral motivation, their average productivity is high; when α is low, on the other hand, the number of green firms is large, but their average productivity is lower. A similar reasoning holds for the brown firms: If α is high, there are many brown firms, and their average productivity is relatively good, but if α is low, there are few brown firms, employing the least effective workers of all.

Let $\Delta w(\alpha)$ denote the *maximum wage difference*, i.e. the maximum extra wage a brown firm is able to offer as compared to a green firm, given the threshold α. If green firms can in fact offer higher wages than brown firms – which can occur, due to the green firms' more productive workers – the maximum wage difference is negative.[13][14]

Let individuals be sorted by their α_i, so that we always have $\alpha_i \leq \alpha_{i+1}$. We are now ready to state our main results.

Proposition 2 *For a separating equilibrium to exist (i.e. there are both green and brown firms in equilibrium), there must exist an individual m and an individual $m+1$ such that*

$$\phi(\alpha_{m+1}) \geq \Delta w(\alpha_m) \geq \phi(\alpha_m) \tag{19}$$

[12] If $\alpha = 0$ no worker prefers to work in a green firm, and if $\alpha > \bar{\alpha} = \max_i \alpha_i$ then nobody prefers to work in a green firm. Note that if individual i chooses green, so does every j for whom $\alpha_j > \alpha_i$; and if i chooses brown, so does every j with $\alpha_j < \alpha_i$. The distribution of α_i is kept fixed throughout the discussion below.

[13] In the case where no brown firm exists, we define $\Delta w(\alpha)$ as the maximum extra wage an entrying brown firm *would be* able to offer, provided that it would only be able to hire workers with $\alpha_i = 0$. $\Delta w(\alpha)$ is similarly defined for $\alpha = \bar{\alpha}$.

[14] Note also that $\Delta w(\alpha)$ is not continuous with a discrete labour force, and while $\phi(\alpha)$ is continuous, α_i only takes a finite number of values. The screening equilibrium may thus fail to exist, even when the smoothed curves $\Delta w(\alpha)$ and $\phi(\alpha)$ intersects. A formal solution to the existence problem would be to assume that firms are uncertain about values of α_i , and that the probability distribution of α_i is continuous. To simplify the analysis we have ignored this problem, which, after all, is similar to the lack of equilibrium in any market with discrete prices and quantities.

where m is the marginal worker employed by a brown firm. If no such worker exists, the equilibrium will be a pooling equilibrium with only green or only brown firms. If $\phi(\alpha_i) > \Delta w(\alpha_i)$ for all α_i, then all firms will be green, while if $\phi(\alpha_i) < \Delta w(\alpha_i)$ for all α_i, all firms will be brown.

Proof. See Appendix. ■

While willingness to pay is unambiguously increasing in α_i, as shown in Lemma 4, note that the slope of $\Delta w(\alpha)$ may be positive or negative. The figures below are thus based on a numerical example (details can be found in the Appendix). Figure 1 depicts workers' willingness to pay (the thick broken line) as a function of their α_i, whereas the solid line illustrates the maximum wage difference when $\alpha_i = \alpha$. Here, all i with $\alpha_i > \alpha^*$ will be employed by green firms (in this case, about half the labour stock), while the rest are employed by brown firms.

Figure 1 about here

It is interesting to compare this to an alternative model without screening of workers, in which individual effort and willingness to pay were uncorrelated[15]. While green firms still pay slightly lower wages, they no longer have more productive workers, so being green pays only if willingness to pay is quite high. In Figure 1, the thin broken line shows the maximum wage difference $\Delta w(\alpha_i)$ in such a model. Without screening, brown firms would be able to offer a wage exceeding every worker's willingness to pay; hence no green firms could exist in equilibrium. With screening, however, the maximum wage difference is much smaller, and we get a separating equilibrium with both types of firms.

A common claim is that due to their abatement costs, green firms will be driven out of business by less altruistic competitors. Here, this may in fact be turned upside down: If $\phi(\alpha_i) > \Delta w(\alpha_i)$ for all α_i, it is the *brown* firms that are driven out of business. The following shows that this is not an empty condition.

Proposition 3 *For other parameters fixed, if there are individuals i, j such that $\alpha_i \neq \alpha_j$ then there exists a*

[15] See Appendix B.1 for details.

constant $\underline{A} > 0$, *such that*

$$\phi(\alpha) > \Delta w(\alpha) \quad \textit{for all } \alpha \in [0, \bar{\alpha}] \quad \textit{if } A < \underline{A}$$

Proof. See appendix. ■

The proposition demonstrates that if abatement costs are low (but strictly positive), then *brown* firms are driven out of business. In this case, green firms can offer a higher wage than any potential brown entrant, since the workers of green firms are, on average, more productive than those an entrying brown firm would be able to recruit[16]. Note that the assumptions of Proposition 3 do not rule out the possibility that a substantial part of the labour force has no moral motivation, $\alpha_i = 0$; the only requirement is that there is at least some heterogeneity ($\alpha_i \neq \alpha_j$). Thus, if abatement is inexpensive, the importance of screening outweights the cost of abatement, and brown firms will be driven out of business.

Screening, which is crucial for the above result, is more important the larger effect effort has on productivity. In our model, this is measured by the parameter μ. In fact, a shift in this parameter can be sufficient to move the economy from an initial situation with no green firms to another with *only* green firms.

Figure 2 about here

This is illustrated in Figure 2a and b, where all parameter values are kept as in Figure 1, exept μ, which is *lower* in Figure 2a and *higher* in Figure 2b (see Appendix for details). In the case depicted in Figure 2a, willingness to pay falls short of the maximum wage difference for all workers, so there are no green firms in equilibrium. Then, μ increases, creating the situation shown in Figure 2b with no brown firm in equilibrium.

Since each worker with $e_i > 0$ becomes more productive than before when μ increases, willingness to pay to work in a green firm actually becomes higher (because in brown firms, higher productivity induce a larger environmental externality). However, for small α the willingness to pay is still small, so the most important effect is on the maximum wage difference curve: When worker effort is very important for productivity, getting highly motivated workers is crucial. Consequently, in the new equilibrium, there are no brown firms

[16] To see this, note that $\phi(0) = 0$. The proposition then implies that in the case with low abatement cost ($A < \underline{A}$), we have $\Delta w(0) < \phi(0) = 0$.

at all. If a brown entrant should emerge, it would only recruit the very least motivated workers, and thus not be able to survive.[17]

7 Conclusions

Our analysis has demonstrated that in a perfectly competitive market, green firms may be able to survive in the long run. The social corporate responsibility profile of a firm may affect both wages and worker productivity. We have demonstrated how this can occur through three channels: First, morally motivated workers demand lower wages from a green than a brown firm. Second, morally motivated workers exert higher effort if employed in a green rather than a brown firm. And, last but not least, a high level of corporate social responsibility, combined with a low wage, can work as a screening device to attract more productive workers.

Two crucial assumptions behind these results are that workers are motivated by a common ethical principle affecting various aspects of their behavior, and, moreover, that the strength of such moral motivation differs between workers. Hence, a given worker's behavior in one context will be correlated with his behavior in other contexts. This correlation is the driving force of the mechanisms described in our paper, allowing firms to use their social responsibility profile for screening purposes. We have demonstrated that this screening may even be powerful enough to drive all brown firms out of business, even when a substantial proportion of workers have no moral motivation at all.

One should not, however, draw the conclusion that workers' moral motivation provides an easy and satisfactory solution to society's environmental and/or moral hazard problems. In the formal model above, green firms were assumed to have no environmentally damaging emissions at all, but in practice there may of course be damages even from green firms. Moreover, although moral motivation does to some extent internalize external effects, our model does not ensure that the market equilibrium is socially optimal. In

[17] Due to the complexity of the model we have not been able to prove as a general result that increasing μ *always* (weakly) increases the share of green firms.

fact, within a slightly different model of moral motivation, Brekke et al. (2003) demonstrated that in general there is undersupply of public goods in equilibrium, even if moral motivation is strong: As the public good is approaching its socially optimal level, the marginal self-image improvement of contributing decreases, while the individual cost of contributing does not. Hence, individuals will contribute more than in a traditional model, but still less than the socially optimal level.

How would public policy measures like green taxes perform in a market like the one considered above? Consider the introduction of a pollution tax in a market where some, but not all firms are green. The effect of this may in fact depend on how workers perceive this tax. If they believe that the tax is insufficiently low, the tax may work more or less as in standard neoclassical models. However, if workes perceive the tax as a correctly set Pigou tax, internalizing external effects completely, there will be no need for the workers to take an individual responsibility for the firm's pollution: If the firm compensates society sufficiently for the damages it causes, both types of firms will be perceived as equally socially beneficial. The implication is that a green tax can in fact crowd out moral motivation completely. If the tax is indeed an adequate Pigou tax, however, there is no real need for workers to take responsibility anyway – so the latter is not necessarily a problem from a social welfare point of view. It *would* be a problem, however, if workers *think* that the tax is a sufficient Pigou tax, while it is, in fact, insufficient.

Our analysis assumes that all firms are identical except for their choice of greenness. However, if firms differ in the sense that non-observable effort is more important for some firms than for others, these firms may be more likely to use a high corporate social responsibility profile as a screening mechanism to attract reliable workers. This would be an interesting extension to explore in future research.

Another possible extention of the model is to look further into the issue of fairness. In our model it is assumed that capital costs are fixed, and that increased productivity benefits all workers in the long run, through higher wages. However, assume that workers believe, instead, that if productivity increases, capital owners (or CEOs) will reap the gains for themselves. Assume, furthermore, that they consider this unfair. In this case, the question "what would happen if every worker acted like me?" would have a very different answer, and within the logic of our model, this could seriously undermine workers' work morale.

This may provide one possible explanation of the findings of Fehr and Falk (2002) and others, indicating that perceptions about the employer's good *intentions* towards the worker is important for workers' willingness to exert unobservable effort.

References

[1] Arora, S., and S. Gangopadhyay (1995): Toward a Theoretical Model of Voluntary Overcompliance, *Journal of Economic Behavior and Organization* **28**, 289-309.

[2] Audi, R. (ed.) (1995): *The Cambridge Dictionary of Philosophy,* Cambridge, UK: Cambridge University Press.

[3] Björner, T.B., L.G. Hansen, and C.S. Russell (2004): Environmental Labeling and Consumers' Choice – an Experimental Analysis of the Effect of the Nordic Swan, *Journal of Environmental Economics and Management* **47** (3), 411-434.

[4] Bolton, G.E. and A. Ockenfels (2000): ERC: A Theory of Equity, Reciprocity, and Competition, American Economic Review, 92 (1), 166-193.

[5] Brekke, K. A., Lurås, H., Nyborg, K. (1996): Allowing Disagreement in Evaluations of Social Welfare, Journal of Economics 63, 303-324.

[6] Brekke, K. A., S. Kverndokk, and K. Nyborg (2003): An Economic Model of Moral Motivation, Journal of Public Economics 87 (9-10), 1967-1983.

[7] Bruvoll, A., B. Halvorsen, and K. Nyborg (2002): Households' Recycling Efforts, Resources, Conservation & Recycling 36 (4), 337-354.

[8] Cropper, M. L., and W. E. Oates (1992): Environmental Economics: A Survey, Journal of Economic Literature 30 (2), 675-740.

[9] European Union Commission (2002): Communication from the Commission concerning Corporate Social Responsibility: A business contribution to Sustainable Development. COM(2002) 347 final. Available at http://europa.eu.int/eur-lex/pri/en/dpi/cnc/doc/2002/com2002_0347en01.doc.

[10] Fehr, E., and A. Falk (2002): Psychological Foundations of Incentives, European Economic Review 46, 687-724.

[11] Fehr, E. and K. M. Schmidt (1999): A Theory of Fairness, Competition, and Cooperation, Quarterly Journal of Economics 114, 817-868.

[12] Frank, Robert H. (1988) *Passion within Reason; the Strategic Role of the Emotions*, Norton.

[13] Frank, Robert H. (2003): What Price the Moral High Ground? Ethical Dilemmas in Competitive Environments, Princeton University Press.

[14] Holmstrom, B. (1982): Moral Hazard in Teams, Bell Journal of Economics 13, 324-340.

[15] Johansson, P.O. (1992): Altruism in Cost-Benefit Analysis, Environmental and Resource Economics 2(6), 605-613.

[16] Maxwell, J.W., T. P. Lyon, and S.C. Hackett (2000): Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism. *Journal of Law and Economics* **43** (2), 583-618.

[17] Holmstrom, B., and P. Milgrom (1991): Multi-Task Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design, Journal of Law, Economics & Organization 7, 24-51.

[18] Milgrom, P. (1993): Is Sympathy an Economic Value? Philosophy, Economics, and the Contingent Valuation Method, in J.A. Hausman (ed.): Contingent Valuation: A Critical Assessment, New York: North-Holland, 417-435.

[19] Moon, W., W.J. Florkowski, B. Bruckner, and I. Schonhof (2002): Willingness to Pay for Environmental Practices: Implications for Eco-Labeling, *Land Economics* **78** (1), 88-102.

[20] Nyborg, K. (2000): Homo Economicus and Homo Politicus: Interpretation and Aggregation of Environmental Values, Journal of Economic Behavior and Organization 42/3, 305-322.

[21] Nyborg, K., and M. Rege (2003): Does Public Policy Crowd Out Private Contributions to Public Goods? Public Choice 115 (3): 397-418.

[22] Schram, A. (2000): Sorting Out the Seeking: The Economics of Individual Motivations, Public Choice 103 (3-4), 231-58.

[23] Spence, A. M. (1973): Job Market Signaling. Quarterly Journal of Economics 87(3), 355-374.

[24] Stiglitz, J. E. (1975): The Theory of Screening, Education, and the Distribution of Income, *American Economic Review* **65** (3), 283-300.

[25] Sugden, R. (1984): Reciprocity: The Supply of Public Goods through Voluntary Contributions, The Economic Journal 94, 772-787.

A Proofs of Propositions and Lemmas

Proof. (Lemma 1):

$\widetilde{V}(e_i, \tau_i)$ is defined as social welfare if everybody made the same choices as i. Acknowledging that individual income and environmental quality may depend on others' behavior, we can formalize this as

$$\widetilde{V}(e_i, \tau_i) = \sum_{j=1}^{n} (u(x_j(e_i, \tau_i), E(e_i, \tau_i), e_j) + S_j) \text{ s.t. } e_j = e_i \text{ and } \tau_j = \tau_i \text{ for all } j.$$

Let $\widetilde{Y}(e_i, \tau_i)$ denote the value of any variable Y in the hypothetical case that $e_i = e_j$ and $\tau_j = \tau_i$ for all $i, j \in \{1, ..., N\}$. Inserting from (4), using that $e_j = e_i$ and $\tau_j = \tau_i$ for all j, and rearranging, then gives

$$\widetilde{V}(e_i, \tau_i) - \sum_{j=1}^{n} \alpha_j \widetilde{V}(e_i, \tau_i) = \sum_{j=1}^{n} (u(\tilde{x}_i(e_i, \tau_i), \tilde{E}(e_i, \tau_i), e_i))$$

Since everybody acts like i, has the same "material" utility function $u(\cdot)$, and agrees on the social welfare function, both e_i and $\widetilde{V}(e_i, \tau_i)$ must be the same for all. This implies that

$$\widetilde{V}(e_i, \tau_i) = NK(u(\tilde{x}_i(e_i, \tau_i), \tilde{E}(e_i, \tau_i), e_i)) \tag{20}$$

where $K = 1/[1 - \sum_{j=1}^{n} \alpha_j]$. Using equations (1) and (4), we then have

$$U_i = u(x_i, E, e_i) + \alpha_i NK(u(\tilde{x}_i(e_i, \tau_i), \tilde{E}(e_i, \tau_i), e_i)).$$

■

Proof. (Lemma 2):

Combining (8) - (10), we get

$$[f(N^h(1 + \mu\bar{e}^h)) - F - \tau^h A]/N^h = f'(N^h(1 + \mu\bar{e}^h))(1 + \mu\bar{e}^h)$$

Using (6), this can be rewritten as

$$[f(L^h) - F - \tau^h A]/L^h = f'(L^h) \tag{21}$$

Monotonicity and concavity of f implies that, given τ^h, this equation has a unique solution for L^h:

$$L^h = L(\tau^h) \tag{22}$$

Inserting this in (10), using (6), and rearranging, we get

$$w(\tau^h) = f'(L(\tau^h))(1 + \mu\bar{e}^h).$$

■

Proof. (Lemma 3):

Proof. Differentiation of (20) with respect to e_i, taking τ_i as given, yields

$$\partial\widetilde{V}(e_i, \tau_i)/\partial e_i = NK[u'_x(\partial\widetilde{x}/\partial e_i) + u'_E(\partial\widetilde{E}/\partial e_i) + u'_e)] \tag{23}$$

Let us calculate $\partial\widetilde{x}/\partial e_i$ and $u'_E\partial\widetilde{E}/\partial e_i$ separately. (10), (6), (22), and the fact that by assumption, $\tau^h = \tau_i$, for all h, we have

$$\widetilde{x}(e_i, \tau_i) = (1 + \mu e_i)f'(L(\tau_i))$$

Differentiating, we get

$$\partial \widetilde{x}(e_i, \tau_i)/\partial e_i = \mu f'(L(\tau_i)) \tag{24}$$

Further, using (12); ■

$$\widetilde{E}(e_i, \tau_i) = E^0 - z[1-\tau_i]H(f(N^h(1+\mu e_i))$$

The number of firms, H, is endogenous, but when τ_i is the same for everyone, all firms will have the same number of workers. Hence, in this case,

$$H = N/N^h = N(\frac{(1+\mu e_i)}{L(\tau_i)})$$

and we can write

$$\widetilde{E}(e_i, \tau_i) = E^0 - [z(1-\tau_i)N\frac{(1+\mu e_i)}{L(\tau_i)}f(L(\tau_i))]$$

The term in square brackets above always equals zero when $\tau_i = 1$, hence we can replace $L(\tau_i)$ in this expression with $L(0)$ (effective labor input per brown firm). Differentiating with respect to e_i then gives

$$\partial \widetilde{E}(e_i, \tau_i)/\partial e_i = -[1-\tau_i]\mu z f(L(0))\frac{N}{L(0)}. \tag{25}$$

Inserting (24) and (25) in (23) and rearranging gives (13), which is the statement to be proved. ■

Proof. (Proposition 1): Differentiation of (16) gives

$$\frac{de_i}{d\alpha_i} = \frac{NK(\mu u'_x f' - \mu u'_E([1-\tau_i]zf(L(0))\frac{N}{L(0)} + u'_e)}{-u''_{ee}(1+NK\alpha_i)}$$

The denominator is strictly positive. Hence, if the numerator is strictly positive, effort is increasing in α_i.

When $\tau_i = 1$, we have that

$$\frac{de_i}{d\alpha_i} = \frac{NK(\mu u'_x f' + u'_e)}{-u''_{ee}(1+NK\alpha_i)} > 0.$$

To see why the inequality hold, note that the first order conditition

$$\alpha_i \partial \widetilde{V}(e_i, \tau_i)/\partial e_i = -u_e$$

have an interior solution since $u'_e(0) = 0$, and hence $\mu f' u'_x + u'_e = \frac{-u'_e}{NK\alpha_i} > 0$ for $\alpha_i > 0$.

When $\tau_i = 0$, we have that

$$\frac{de_i}{d\alpha_i} = \frac{NK(\mu u'_x f' - \mu u'_E (zf(L(0))\frac{N}{L(0)}) + u'_e)}{-u''_{ee}(1 + NK\alpha_i)}.$$

Now, if $\partial\widetilde{V}(0,0)/\partial e_i \leq 0$, $\alpha_i V(e_i, 0) + \bar{u}(e_i)$ obviously have its maximum at $e_i = 0$, (since $e_i \geq 0$). Otherwise, with $\partial\widetilde{V}(0,0)/\partial e_i > 0$, the first order condition have an interior solution, and hence

$$\mu u'_x f - \mu u'_E (zf(L(0))\frac{N}{L(0)}) + u'_e = \frac{-u'_e}{NK\alpha_i} > 0$$

and $\frac{de_i}{d\alpha_i} > 0$.

From the first order condtion

$$\alpha_i \partial\widetilde{V}(e_i, \tau_i)/\partial e_i = -u_e$$

the claim that effort is higher in green firm is a direct consequence of axiom 1, where we assume that

$$\partial\widetilde{V}(e_i, 1)/\partial e_i > \partial\widetilde{V}(e_i, 0)/\partial e_i.$$

. ■

Proof. (Lemma 4): By the definition of willingness to pay (17),

$$u(w(0) - \phi_i, E, e_i^1) + \alpha_i(\widetilde{V}(e_i^1, 1)) = u(w(0), E, e_i^0) + \alpha_i(\widetilde{V}(e_i^0, 0)).$$

Proposition 1 assumes that $u(x_i, E, e_i) = \widehat{u}(x_i, E) + \overline{u}(e_i)$, where $\widehat{u}'_x > 0, \widehat{u}'_E > 0, \overline{u}' < 0$ and $\overline{u}'' < 0$. Inserting this and rearranging gives

$$\widehat{u}(w(0) - \phi_i, E) - \widehat{u}(w(0), E) = \alpha_i[\widetilde{V}(e_i^0, 0) - \widetilde{V}(e_i^1, 1)] + \overline{u}(e_i^0) - \overline{u}(e_i^1)$$

We differentiate this implicitly with respect to α_i :

$$\begin{aligned} -\widehat{u}'_x(w(0) - \phi_i, E)\frac{d\phi_i}{d\alpha_i} &= [\widetilde{V}(e_i^0, 0) - \widetilde{V}(e_i^1, 1)] + \alpha_i \frac{d}{d\alpha_i}[\widetilde{V}(e_i^0, 0) - \widetilde{V}(e_i^1, 1)] + \overline{u}'\frac{de_i^0}{d\alpha_i} - \overline{u}'\frac{de_i^1}{d\alpha_i} \\ &= [\widetilde{V}(e_i^0, 0) - \widetilde{V}(e_i^1, 1)] + \frac{de_i^0}{d\alpha_i}\left[\overline{u}'(e_i^0) + \alpha_i \widetilde{V}'_e(e_i^0, 0)\right] - \frac{de_i^1}{d\alpha_i}\left[\overline{u}'(e_i^1) + \widetilde{V}(e_i^1, 0)\right] \end{aligned}$$

The two last terms on the right hand side both equal zero, due to the first order condition for effort. Hence

$$\frac{d\phi_i}{d\alpha_i} = -\frac{\widetilde{V}(e_i^0, 0) - \widetilde{V}(e_i^1, 1)}{\widehat{u}'_x(w(0) - \phi_i, E)}$$

From (14), we know that $\widetilde{V}(e_i, 1) > \widetilde{V}(e_i, 0)$ for any given e_i. However, here we may have that $e_i^0 \neq e_i^1$, since workers exert higher effort in green firms. Effort has a social cost. Nevertheless, we know that $e_i^0 < e_i^1$ *only* if this is socially optimal, i.e. only if $\widetilde{V}(e_i^1, 1) > \widetilde{V}(e_i^0, 1)$; hence

$$\widetilde{V}(e_i^G, 1) > \widetilde{V}(e_i^B, 1) > \widetilde{V}(e_i^B, 0),$$

and

$$\frac{d\phi_i}{d\alpha_i} > 0.$$

■

Proof. (Proposition 2): For a separating equilibrium to exist, there must a) exist a worker $m+1$ who prefers green firms in equilibrium, and b) another worker m who prefers brown firms in equilibrium, with $\alpha_{m+1} \geq \alpha_m$. The first equality in (19), $\phi(\alpha_{m+1}) \geq \Delta w(\alpha_m)$, ensures a), due to (18). Since workers disregard their own impact on equilibrium wages, the fact that we could have $\Delta w(\alpha_m) \neq \Delta w(\alpha_{m+1})$ will not affect behavior. The second inequality in (19), $\Delta w(\alpha_m) \geq \phi(\alpha_m)$, ensures b).This follows immediately from the definition of ϕ_i . If $\phi(\alpha_i) > \Delta w(\alpha_i)$ for all α_i, then no matter how the market is divided into green and brown firms, the marginal worker will always strictly prefer green. The proof of the reverse inquality is similar. ■

Before proving Proposition **3,** we first need a lemma.

Lemma 5 *For a fixed α, $\phi(\alpha)$ is decreasing in A, while $\Delta w(\alpha)$ is increasing in A.*

Proof. From the definition of ϕ

$$u(w(0) - \phi(\alpha_i), E, e_i^1) + \alpha_i(\widetilde{V}(e_i^1, 1)) = u(w(0), E, e_i^0) + \alpha_i(\widetilde{V}(e_i^0, 0))$$

Now, a positive shift in A does not affect the technology of the brown firm, and hence the right hand side is unaffected by the shift. On the left hand side, $u(w(0) - \phi(\alpha_i), E, e_i^1) + \alpha_i(\widetilde{V}(e_i^1, 1))$ changes due to changes in e_i^1 which can be ignored due to the first order condition for optimal effort e_i^1. In addition there is a direct effect $\partial\tilde{V}/\partial A > 0$ since abatement is more costly, and hence social welfare decreases. It follows that $\phi(\alpha_i)$ must decrease to offset the shift in $\tilde{V}$.

For $\Delta w(\alpha)$, we note that wages in the brown firm are unaffected by changes in A, while by the definition (8), W^1 must be decreasing in A. ■

Proof. (Proposition 3): Note first that with $A = 0$, there is no cost of abatement, while the selection makes the green firms most productive, and hence $\Delta w(\alpha) < 0$ while $\phi(\alpha) > 0$. Next, as $A \to A^M$, the wage differences will increase while $\phi(\alpha) \to 0$. Now if $\Delta w(\alpha)$ is strictly positive for some $A < A^M$, let $A^*(\alpha)$ be such that $\phi(\alpha) = \Delta w(\alpha)$. If $\Delta w(\alpha) \leq 0$ for all $A < A^M$ then $A^*(\alpha) = A^M$. Now, $A^*(\alpha)$ must be continuous, since both $\Delta w(\alpha)$ and $\Delta w(\alpha)$ are continuous functions of A. As $\Delta w(\alpha) < 0\phi(\alpha)$, for $A = 0$ it further follows that $A^*(\alpha)$ must be bounded away from zero. It follows that $A^*(\alpha)$ will exhibit a strictly postive minimum on a closed interval containing the support of α_i. ■

B A worked example

Our figures are based on the following example. Let

$$f(L) = L^{\gamma}.$$

Inserting into (21) yields

$$\begin{aligned} L(1) &= (\frac{F+A}{1-\gamma})^{1/\gamma} \\ L(0) &= (\frac{F}{1-\gamma})^{1/\gamma} \end{aligned}$$

Thus, wages in the two firms are

$$\begin{aligned} w(0) &= f'(L(0))(1+\mu\bar{e}^B) = \gamma(\frac{F}{1-\gamma})^{(\gamma-1)/\gamma}(1+\mu\bar{e}^B) \\ w(1) &= f'(L(1))(1+\mu\bar{e}^G) = \gamma(\frac{F+A}{1-\gamma})^{(\gamma-1)/\gamma}(1+\mu\bar{e}^G) \end{aligned}$$

Next, let

$$u(x, E, e) = x + \beta E - \frac{1}{b}e^b \text{ with } b > 1$$

Now,

$$
\begin{aligned}
\phi(\alpha) &= \alpha\left[\widetilde{V}(e^G(\alpha),1)-\widetilde{V}(e^B(\alpha),0)\right] \\
&= \alpha NK\left[f'(L(1))(1+\mu e^G(\alpha))-f'(L(0))(1+e^B(\alpha))+\beta(1+e^B(\alpha))zf(L(0))\frac{N}{L(0)}\right] \\
&= \alpha NK\left[\gamma(\frac{F+A}{1-\gamma})^{(\gamma-1)/\gamma}(1+\mu e^G(\alpha))-(\gamma-\beta zN)(\frac{F}{1-\gamma})^{(\gamma-1)/\gamma}(1+e^B(\alpha))\right]
\end{aligned}
$$

Since we assume that effort enhances self-image even in brown firms, it follows that $\gamma > \beta zN$.

From (16) we find that

$$
\begin{aligned}
e^G(\alpha) &= \mu^{b/(b-1)}\left[\frac{\alpha NK}{1+\alpha NK}\left(f'(L(1))\right)\right]^{1/(b-1)} = \mu^{b/(b-1)}\left[\frac{\alpha NK}{1+\alpha NK}\gamma(\frac{F+A}{1-\gamma})^{(\gamma-1)/\gamma}\right]^{1/(b-1)} \\
e^B(\alpha) &= \mu^{b/(b-1)}\left[\frac{\alpha NK}{1+\alpha NK}\left(f'(L(0)-\beta zf(L(0))\frac{N}{L(0)}\right)\right]^{1/(b-1)} \\
&= \mu^{b/(b-1)}\left[\frac{\alpha NK}{1+\alpha NK}\left((\gamma-\beta zN)(\frac{F}{1-\gamma})^{(\gamma-1)/\gamma}\right)\right]^{1/(b-1)}
\end{aligned}
$$

Given these effort levels, we can compute $\bar{e}$ for the two types of firm and different values of α. In the simulations we assume $NK = 200, \gamma = 0.5, \beta zN = 0.3$, and $b = 1.5$. In figure 1, $\mu = 3$, in Figure 2(a) $\mu = 2$ and in Figure 2(b) $\mu = 4$. Wages and willingness to pay are measured in dollars.

B.1 No screening

The case with no screening in Figure 1 assumes that the self-image function (4) is replaced by

$$S_i = \delta_i\psi(e) + \alpha_i\sigma(\tau_i)$$

where ψ and σ are functions which are weighted by weights δ_i and α_i. The weight α_i is as before while δ_i is identically distributed but independent of α_i. All other functional forms and parameter values are kept as in the worked example above. In Figure 1 we have, for comparison, calibrated the self-image functions such that effort as a function of δ_i corresponds to effort in green firms as a function of α_i, that is, an individual with effort weight δ_i in the alternative model, and an individual working in a green firm and with moral motivation $\alpha_i = \delta_i$ in the original model, will both provide the exact same effort. Similarly the willingness to pay function $\phi(\alpha_i)$ is the same in both models. In this case, workers' willingness to pay to work in green firms would be unchanged; but, although workers may also take pride in exerting more effort than the minimum level, willingness to pay and effort are no longer correlated, so in such a model there could be no screening.

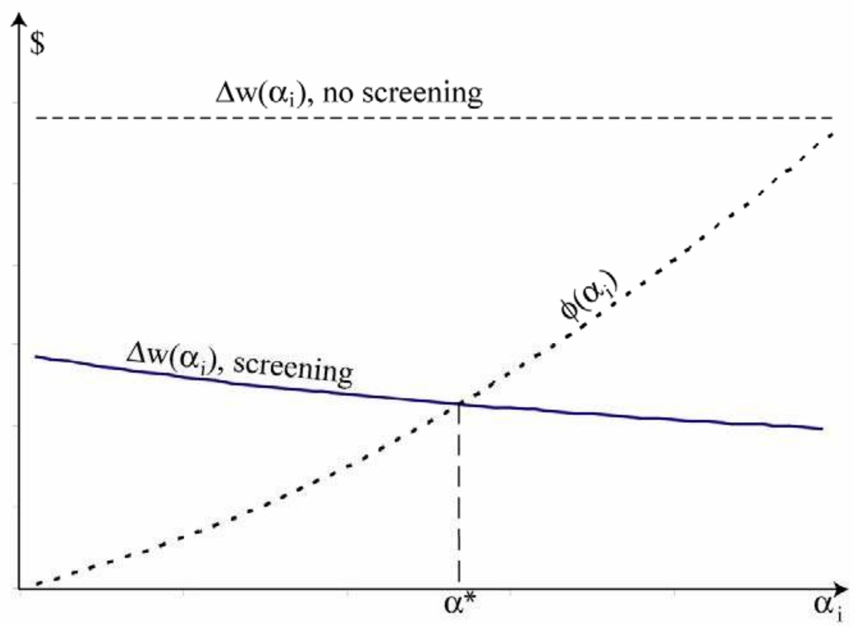

Figure 1: Wage difference and willingness to pay, with and without screening.

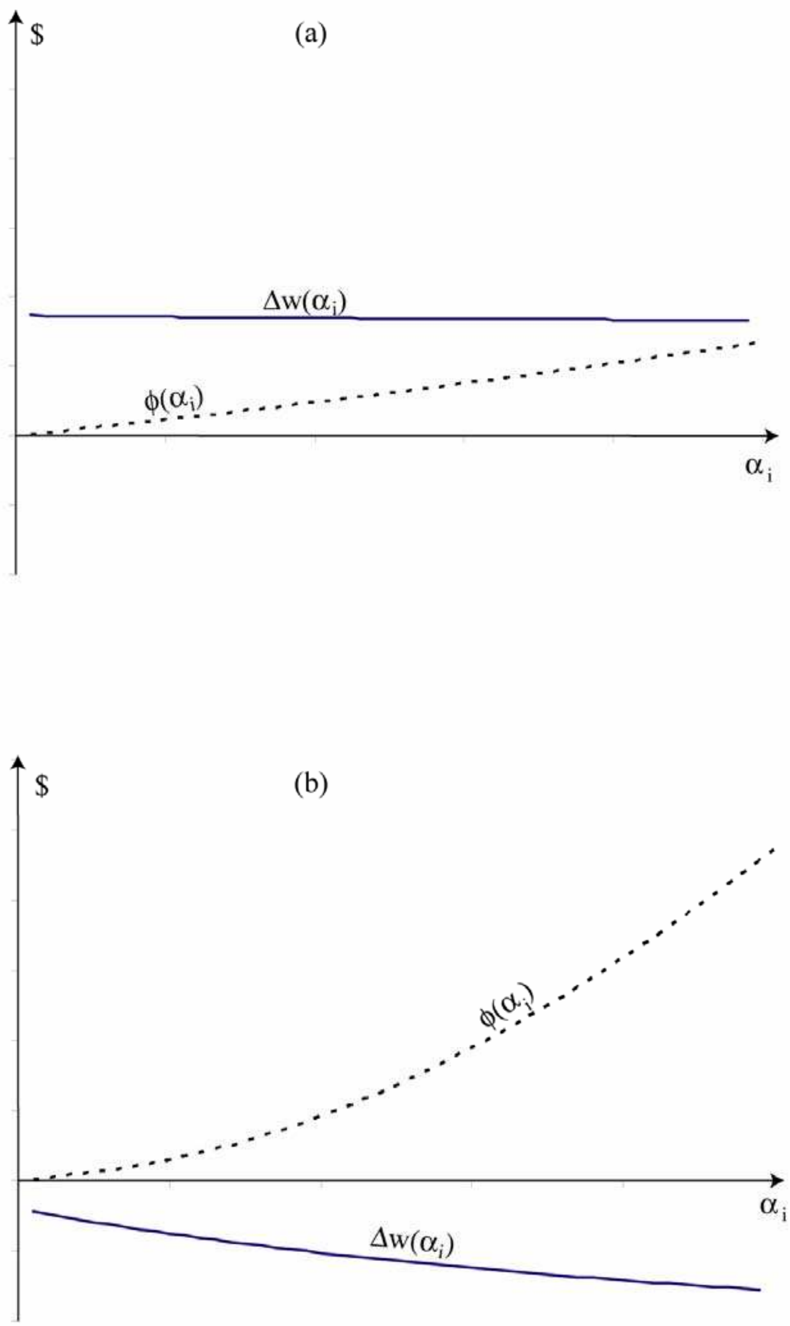

Figure 2: Willingness to pay and wage difference for low (a) and high (b) values of μ.