

---

## **Attracting responsible employees: Green production as labor market screening**

Kjell Arne Brekke and Karine Nyborg

Addresses:

Department of Economics, University of Oslo, P.O.Box 1095 Blindern, NO-0317 Oslo, Norway (both authors). Nyborg (corresponding author): karine.nyborg@econ.uio.no, ph. +4722857283, fax +4722855035. Brekke: k.a.brekke@econ.uio.no, ph. +4722841169, fax +4722855035.

Acknowledgements: Funding from the Research Council of Norway through the SAMSTEMT program is gratefully acknowledged. This paper was written while both authors were employed by the Ragnar Frisch Centre for Economic Research; part of the research was also conducted while Brekke was employed by the Center for Development and the Environment, University of Oslo. Thanks to numerous seminar and conference participants for helpful and encouraging comments.

# Attracting responsible employees: Green production as labor market screening

September 12, 2007

## Abstract

Corporate social responsibility can improve firms' ability to recruit highly motivated employees. This can secure socially responsible firms' survival even in a highly competitive environment. We show that if both socially responsible (green) and non-responsible (brown) firms exist in equilibrium, workers with high moral motivation, who shirk less than others, will self-select into the green firms. If unobservable effort is sufficiently important for firm productivity, this can drive every brown firm out of business – even in the case where many workers have no moral motivation whatsoever.

*Keywords:* Moral hazard, moral motivation, teamwork, corporate social responsibility, voluntary abatement.

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

## 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 not mandated by law, and commit to ethical principles increasing their production costs, such as abstaining from the use of child labor in developing countries.<sup>1</sup> But why would a private firm pay extra costs to promote social values? If it

---

<sup>1</sup>For example, Exxon, Chiquita, McDonald's, Coca-Cola and Ford Motor Company all have information concerning their corporate social responsibility commitments figuring prominently on their homepages (January 2007), together

does, will it not be wiped out of business by less responsible competitors?

Corporate social responsibility is often explained by consumers' willingness to pay for ethically produced goods (Arora and Gangopadhyay 1995, Björner et al. 2004); or by investors' and/or entrepreneurs' willingness to abstain from maximum profits to promote social values (Cullis et al. 1992, Baron 2005). Other proposed explanations include pre-emption of new taxes or regulations (Maxwell et al. 2000), or signalling a "nice type" in order to get lenient treatment when new regulations are about to be introduced (Heyes 2005a). While we acknowledge the relevance and importance of these explanations, we will disregard them below. Our focus will be, instead, on an issue hardly touched upon in the economics literature on corporate social responsibility, namely employee motivation.<sup>2</sup> The basic idea is simple: responsible employers attract responsible workers, who shirk less.

For brevity, a socially responsible firm will be termed "green" below, while other firms are called "brown". However, the main idea applies for corporate social responsibility in general, not just its environmental aspect. Let us explain the intuition of our argument in three steps:

First, if some workers strictly prefer green employers, all else given, this may result in a lower equilibrium wage in green than in brown firms. Such a wage differential will keep green production in business, provided that green firms' extra cleaning costs are offset by their lower wage payments. This is a very straightforward and almost trivial point, hardly requiring formal proof.<sup>3</sup> Nevertheless, the problem with this explanation is that workers' willingness to pay for green employment would often have to be substantial: For green firms to be equally profitable as brown ones, workers in green firms essentially would have to share the entire cleaning cost between them. In practice, willingness to pay for green employment might not be that high.

Assume, now, that individual effort is imperfectly observable. The second step in our argument is the following: *If* those workers who prefer green employment are also more likely to work hard, green firms' productivity will be higher. Hence, green firms could survive in market equilibrium, even with reports of costly measures taken to promote social and/or environmental values.

<sup>2</sup>A brief, verbal discussion of CSR and employee motivation can be found in Heal (2005).

<sup>3</sup>Note that while this result holds in a model of identical individuals, step 2 below requires variation in the strength of individuals' moral motivation.

if workers' willingness to pay for green employment is very modest.<sup>4</sup>

Nevertheless, is it reasonable to expect a correlation between who prefers green employment and who works hard? Our third step is to argue that this is, indeed, quite plausible. In fact, we will formalize one specific type of moral preferences, assume that the strength of this preference differs between individuals, and from this derive a perfect correlation between an individual's effort at work and her willingness to pay for green employment.

Our main result depends neither on this correlation to be perfect, nor on its being derived from the particular moral principle that we propose. It does depend, however, on a correlation being present. A person who is very concerned about being responsible, or who enjoys making contributions to social welfare, will presumably derive satisfaction both from contributing to her firm's productivity and to a greener society. Similarly, a person who is much bothered by bad conscience will probably be more reluctant both to shirk and take part in polluting production activities. A person of the Homo Oeconomicus type, on the other hand, will never work in green firms if the pay is lower; and she will definitely shirk if she can. More generally, such correlations arise naturally if moral satisfaction is not derived mechanically from having performing some uni-dimensional action, but is rather related to some morally relevant variable influenced by *many* actions; for example, the social welfare effects of one's behavior. Then, the more an individual cares about being morally responsible, the higher is her concern for all behaviors affecting this morally relevant variable.

Below, we will formalize the above argument, using a specification of moral motivation which is close to that proposed in Brekke et al. (2003). We will demonstrate that if the strength of moral motivation varies between individuals, green firms can survive in equilibrium, even if workers' willingness to pay for green employment is very small. More interestingly, if unobservable effort is sufficiently important for firm productivity, *every* brown firm will be wiped out by competition; even if a substantial share of workers have no moral motivation whatsoever.<sup>5</sup>

---

<sup>4</sup>This logic is closely related to standard screening models (Stiglitz 1975).

<sup>5</sup>We do believe that other types of moral or altruistic motivation can produce similar results as those reported below. However, this would require a separate analysis: Since the relevant types of moral motivation involve endogenous variables such as others' welfare, individual utility maximization cannot be analyzed separately from a description of the rest of the economy. We have made an effort to simplify the model as much as possible; but alternative moral

There is indeed some empirical evidence indicating that many people do prefer their employer to be socially responsible, and that they are willing to accept a lower wage to achieve this. Vitell and Davis (2004) found that job satisfaction was substantially higher when "top management was perceived as strongly supporting ethical behavior" (p.493). Reinikka and Svensson (2003) found that religious non-profit primary health care facilities, providing more services with a public good element and charging lower prices than private for-profit facilities, hired qualified medical staff significantly below the market wage. The employer branding firm Universum collects responses from roughly 180,000 economics, business and technology students in 28 countries in its annual Graduate Survey, and one of their questions is: "Which of the following do you find most important when you select you future ideal employers?" In the 2006 US survey, 21.5 percent of respondents selected "corporate social responsibility", while 39 percent selected "high ethical standards"; the corresponding European averages were 19.9 (CSR) and 21.1 (high ethical standards).<sup>6</sup> Frank (2003), using data for Cornell graduates and controlling for sex, curriculum, and academic performance, found a large and statistically significant compensating salary differential among recent Cornell graduates, with the jobs rated as *less* socially responsible earning substantially *higher* wages.<sup>7</sup>

Concerning the question of whether individual moral choices are correlated, Blanco et al. (2007) studied a sequential Prisoners' Dilemma game and found that when the first mover had chosen to cooperate, the second mover's choice was significantly correlated with the latter's decisions in other games: The choice to cooperate was positively and significantly correlated with this individual's preferences involve different morally relevant variables, implying that the simplifications made in the current paper quickly become too crude.

<sup>6</sup>Personal communication: Carlo Duraturo, March 29, 2007. See also [www.universum.se](http://www.universum.se). The response alternatives included, e.g., "exciting products/services", "financial strength", and "innovation", as well as "corporate social responsibility" and "high ethical standards". Respondents could pick a maximum of 3 items. In US, there were 23 response alternatives; in Europe, the number differed between countries but were between 12 and 23.

<sup>7</sup>Frank also asked survey 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.

contributions in a public good game; with her offers when playing the proposer in an Ultimatum Game; and also with her choice if playing the first mover in a sequential Prisoners' Dilemma game. Altmann et al. (2007), using a modified version of the trust game, found a strong and positive relationship between a person's generosity when playing the first mover and her tendency to repay others' generosity when placed in the role of the second mover.

Although we know of no other paper discussing corporate social responsibility as labor market screening, the problem of identifying and attracting individuals who are particularly highly motivated has been studied by several scholars. For example, Heyes (2005b) points out that increasing wages might attract the 'wrong sort' of people into the nursing profession; Alger and Ma (2003) discuss optimal contracts between an insurer and a provider when the latter may be either of a "truthful" or a "collusive" type; Besley and Ghatak (2005) analyze matching of employers and employees with similar "mission" preferences. Our contribution is to show that when cooperative behavior originates from an underlying general principle of ethics, this affects several aspects of individual behavior in a correlated way, which allows not only for reduced wage payments in green firms, but also for labor market screening. By becoming green, and at the same time reducing wages slightly, a firm will attract only those who have a high moral motivation; and those workers also shirk less. Thus, even if a green firm keeps the same wage as the brown firms, its workers will, on average, be more productive than the workers of the brown firms, because those with high moral motivation will strictly prefer green employment.

## 2 The economy: Production, pollution, and wages

Consider an economy characterized by a large number of profit maximizing firms, a perfectly competitive labor market and full employment. 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 equilibrium, there are no pure rents.

Assume that there is a large number of workers,  $N$ , with identical utility functions. The utility of worker  $i$  is assumed to be increasing in his consumption of private goods  $x_i$  and environmental

quality  $E$ , decreasing in effort  $e_i$ , and, finally, increasing in  $S_i$ , his self-image as a socially responsible individual, in the following way:

$$U_i = x_i - c(e_i) + \gamma E + S_i \quad (1)$$

where  $c'(0) = 0$ ,  $c'(e_i) \geq 0$ ,  $c''(e_i) > 0$  (primes denote derivatives), and  $\gamma > 0$  is the marginal utility of environmental quality. Linear separability and constant marginal utilities of income and environmental quality is assumed for simplicity. We will return to the issue of self-image in the next section.

Production takes place in teams, and individual effort is unobservable. Employers observe the total level of production, but since they cannot distinguish the contributions of each worker, individual wages are equal for all workers within a given firm. Thus, workers have no pecuniary incentive to work hard, and the firm faces a moral hazard problem.<sup>8</sup>

For simplicity, assume that each firm hires exactly  $L$  workers, and that each firm's production is increasing in the average effort exerted by these  $L$  workers. Let  $e^\tau$  denote expected average  $e_i$  among workers in a firm of type  $\tau$ .<sup>9</sup> Let production  $y^\tau$  of a firm of type  $\tau$  be given by

$$y^\tau = (1 + e^\tau)\mu L \quad (2)$$

where  $\mu > 0$ .<sup>10</sup> Firms are assumed to be large enough to make it infeasible for a single worker to notice the change in average productivity resulting from a change in his own individual effort. Workers consequently consider  $e^\tau$  as exogenously given.<sup>11</sup>

---

<sup>8</sup>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.

<sup>9</sup>Below, we will disregard random differences between expected and actual average effort.

<sup>10</sup>Hence, production is strictly positive even if  $e^\tau = 0$ . This can be interpreted as saying that there exists a level of effort below which marginal disutility is strictly negative (the worker becomes bored). Since every worker will exert at least this level of effort, we can use this as the starting point of our effort variable  $e_i$ . Hence,  $e_i$  can be interpreted as worker  $i$ 's voluntary contribution of costly effort.

<sup>11</sup>Parts of our analysis would have been simplified by assuming a continuous population. We have chosen not to do so, partly because this would make our formalization of moral motivation intuitively hard to follow.

Capital costs are fixed and identical for each firm and will be disregarded below. However, each firm emits a fixed amount of hazardous pollution; and end-of-pipe cleaning equipment, eliminating the environmental damage caused by the firm's pollution, is available at a fixed cost  $A$ . This equipment can be purchased and installed by firms on a voluntary basis; no regulation enforcing the use of abatement equipment is assumed to be in place.<sup>12</sup> Consequently, there may potentially exist two types of firms in this economy: Green firms ( $\tau = G$ ) choose to pay  $A$  and do not damage the environment, while brown firms ( $\tau = B$ ) do not pay  $A$ , but do cause environmental damage.

Firms with negative profits cannot survive in the long run. Let  $\pi^\tau$  be the equilibrium profit of a firm of type  $\tau$ . If firm type  $\tau$  exists in equilibrium, we thus have

$$\begin{aligned}\pi^G &= (1 + e^G)\mu L - Lw(G) - A = 0 \\ \pi^B &= (1 + e^B)\mu L - Lw(B) = 0\end{aligned}\tag{3}$$

where  $w(\tau)$  is the wage per worker in firm type  $\tau$ .

Environmental quality, which is a pure public good, is given by an initial level  $E^0$  less the environmental damage caused by pollution, in the following (linear) way:

$$E = E^0 - bZ\tag{4}$$

where  $b \in [0, 1]$  is the share of brown firms, and  $Z > 0$  is the fixed total environmental damage which would result if all firms were brown, i.e. if no firms installed abatement equipment.  $E$  is exogenous in the sense that no single worker is able to notice changes in  $E$  resulting from his own choices; moreover, no worker has reason to believe that his individual choices will influence the share of brown firms in the economy.

A worker  $i$ 's income is given by the wage offered by his employer.<sup>13</sup> Hence, individual  $i$ 's budget constraint is given by

---

<sup>12</sup>Whether abatement eliminates or just reduces environmental damage is not essential. Later, we will relax the assumption that pollution is independent of the firm's production level.

<sup>13</sup>To keep the analysis simple, we will discuss this as if each individual works full-time in one and only one firm. Strictly speaking, the formal analysis below requires that the marginal worker can share his time between two employers.



$$x_i = w(\tau_i) \tag{5}$$

where  $\tau_i \in \{G, B\}$  is the type of the firm  $i$  chooses to work in.

In Brekke and Nyborg (2004), we derive all our main results within a model where workers' utility function is more general, firm size is endogenous, each firm's production function is strictly concave in effective labor input, and pollution is increasing in the firm's production. These generalizations complicate the formal analysis substantially, but – with the exception of the variable pollution assumption – provide little additional insight. To avoid cluttering the analysis, we will thus keep to the simplifications presented above. However, after characterizing long-term equilibrium, we will discuss the implications of production-dependent pollution, since this does provide an interesting additional insight; namely that corporate social responsibility does not only facilitate the recruitment of responsible types, it also increases the work motivation of every given employee with a strictly positive moral motivation.

### 3 Workers' preferences: Morally motivated utility maximizers

Workers maximize their utility by choosing in which firm to seek employment, and, given their employer, how much effort to exert while at work. Before proceeding, however, we need to discuss the issue of self-image and how to formalize this.

Our understanding of "moral motivation" will be closely related to that proposed in Brekke et al. (2003), although the formalization below differs slightly. Individuals are assumed to have preferences for a self-image as socially responsible.<sup>14</sup> To assess his self-image, an individual considers his own actual behavior and then asks himself the hypothetical question: "What would happen to social welfare if everybody acted just like me?" Self-image is determined by the answer to this question: The better the social welfare consequences if everybody had (hypothetically) acted like him, the better is his self-image. This kind of "everyday moral reasoning" can be viewed as inspired by Immanuel Kant's

---

<sup>14</sup>Other papers incorporating concepts of self-image in economic models include, for example, Akerlof and Kranton (2000) and Benabou and Tirole (2006).

categorical imperative: One should act only according to those maxims that can be consistently willed as a universal law (see Audi 1995, p. 403).<sup>15</sup> Furthermore, it is consistent with other 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 (Matthew 7.12). Survey responses indicate that it is indeed common to take such considerations into account; for example, in a Norwegian survey conducted in 1999, 93 percent claimed to recycle at least part of their household waste, and as much as 88 percent of those agreed or agreed partly to the following statement: "I recycle partly because I think I should do what I want others to do" (see Bruvoll et al., 2002).

To judge the social welfare consequences had everybody acted like him, the individual must, of course, have some conception of what social welfare means. Let  $\mathbf{x} = (x_1, \dots, x_N)$  and  $\mathbf{e} = (e_1, \dots, e_N)$ . To make things as simple as possible, assume that every worker has the following utilitarian-type view of social welfare,  $V(\mathbf{x}, \mathbf{e}, E)$ :

$$V(\mathbf{x}, \mathbf{e}, E) = \sum_{j=1}^N [x_j - c(e_j) + \gamma E]. \quad (6)$$

Although the worker considers his own impact on average effort, wage levels, and environmental quality to be negligible, the consequences for these variables could of course not be neglected had everybody behaved just like him. If everybody worked in green firms, for example, environmental damages would be eliminated. Similarly, if everybody increased their effort, this would increase the equilibrium wage, and thus consumption; it would also increase everybody's disutility of effort.

Let  $E(\tau_i)$  denote environmental quality in the hypothetical case that every individual chose the same firm type as  $i$ , that is, if  $\tau_i = \tau_j$  for all  $j$ . Similarly, let  $\mathbf{x}(e_i, \tau_i) = (x_1(e_i, \tau_i), \dots, x_N(e_i, \tau_i)) = (x(e_i, \tau_i), \dots, x(e_i, \tau_i))$  denote the vector of (identical) incomes that would result if  $e_i = e_j$  and  $\tau_i = \tau_j$  for all  $j$ , while  $\mathbf{e}_i$  denotes the situation where everybody exerts the same effort as  $i$ , i.e.  $\mathbf{e}_i = (e_i, \dots, e_i)$ .

We are now ready to specify our self-image function, which says that worker  $i$ 's self-image is

---

<sup>15</sup>The workers modeled here are not Kantians, though, since they are in fact willing to trade the satisfaction of doing the right thing against increased consumption or leisure. A strict Kantian would adhere to the Kantian ethical ideals categorially, not allowing such tradeoffs.

proportional to the social welfare consequences *had everybody made the same choices as him*:

$$S_i = \alpha_i V(\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i)) \quad (7)$$

Here,  $\alpha_i \in [0, \bar{\alpha}]$ , where  $\bar{\alpha} < 1$ , is an individual-specific parameter indicating how important social welfare considerations are for individual  $i$ 's self-image. If  $\alpha_i = 0$ ,  $i$ 's preferences correspond to the traditional Homo Oeconomicus case; if  $\alpha_i$  were equal to 1, which is precluded by assumption, the individual would place just as much emphasis on each single individual's welfare (in the hypothetical situation where everyone acts like himself) as he does on his own actual utility.

Note that, just as the categorical imperative defines one's moral responsibility vis-a-vis society without referring to others' *actual* behavior, there is no presumption in our analysis that the worker thinks others will *in fact* follow his example. When evaluating the moral stance of his action, the worker does not consider the actual impact on social welfare, but the hypothetical impact if his choice was to be made a universal law.<sup>16</sup>

Note also that the social welfare function (6), which is the basis of individuals' self-image considerations, does not include self-image benefits. This can obviously be disputed: On the one hand, it may seem unreasonable to include the benefits of "doing good" in the very definition of "good". On the other hand, it is hard to argue that others' self-image benefits are somehow less "real" than other benefits. However, within our framework, using the social welfare measure (6) for making the self-image evaluations in (7) is *behaviorally equivalent* to assuming a classical all-inclusive utilitarian social welfare function. In the following, we will thus stick to the simple formulation used in equation (6).

**Lemma 1** *Using the social welfare function  $V(\mathbf{x}, \mathbf{e}, E) = \sum_{j=1}^N [x_j - c(e_j) + \gamma E]$  as the basis for self-image evaluations, as specified in equation (7), is behaviorally equivalent to rescaling the parameters  $\alpha_i$  and then using the alternative social welfare function  $V^S = \sum_{j=1}^N U_j$ , provided that, with the latter specification,  $\sum_{j=1}^N \alpha_j < 1$ .*

---

<sup>16</sup>Our formalization here is slightly 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.

**Proof.** See Appendix A. ■

## 4 Workers' effort

Differentiating (1) with respect to  $e_i$ , taking  $\tau_i$  as given and using (5), (6) and (7), yields the following first order condition for an interior utility maximum:

$$\alpha_i dV(\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i))/de_i = c'(e_i) \quad (8)$$

The worker will exert effort until the marginal benefit in terms of a better self-image just equals the marginal disutility of effort. The next question is what determines the left hand side of this equation: How would it affect social welfare if  $i$  worked slightly harder, and everyone followed his example?

This depends on how important average effort is for production, that is, on  $\mu$ , which determines the (potential) effect on wages in long-term equilibrium and thus on everyone's consumption benefits. Differentiating  $V(\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i))$  with respect to  $e_i$ , taking (3) and (5) into account and inserting into the first order condition (8), gives

$$\frac{\alpha_i N \mu}{(1 + \alpha_i N)} = c'(e_i) \quad (9)$$

This condition will hold for all workers, since  $c'(0) = 0$ : Workers with  $\alpha_i = 0$  maximize their utility by providing no costly effort; those with  $\alpha_i > 0$  provide a strictly positive level of costly effort. Note that every 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  $dV[\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i)]/de_i = 0$ , while utility maximization for a worker with  $\alpha_i > 0$  implies  $dV[\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i)]/de_i = c'(e_i)/\alpha_i > 0$ . Hence, although a worker with  $\alpha_i > 0$  does strive towards his conception of a morally ideal behavior, he stops short of reaching that ideal. This implies that voluntary effort will never reach its first-best level.

Now, a crucial point in our argument is that firms want to hire morally motivated workers because these are more productive; they work harder, or equivalently, shirk less. Since (9) holds with equality for all values of  $\alpha_i$ , individual  $i$ 's effort is given by

$$e_i = (c'(e_i))^{-1} \left( \frac{\alpha_i N \mu}{1 + \alpha_i N} \right) \quad (10)$$

The next result follows directly from this:

**Proposition 1** *Worker  $i$ 's effort  $e_i$  is strictly increasing in his moral motivation  $\alpha_i$ .*

Hence, not unexpectedly, moral motivation alleviates the moral hazard problems in team production pointed out by Holmstrom (1982): Highly motivated workers, that is, workers with high values of  $\alpha_i$ , work harder than others, *ceteris paribus*.

Note that for any given worker, effort is independent of the type of firm he works in. If a firm's emissions were increasing in production, any given worker with  $\alpha_i > 0$  would work harder in a green than in a brown firm. This effect would reinforce our main result concerning labor market screening, but is not needed to obtain it. Thus, for the sake of simplicity, we will keep the assumption of fixed pollution per firm, but return to the case of variable emissions later.

## 5 Willingness to pay

Provided that abatement is in fact considered socially preferable to no abatement, working in a green firm provides, all else equal, a higher self-image than working in a brown firm. Morally motivated workers will thus have a strictly positive willingness to pay for working in a green firm. Hence, in equilibrium, green firms may be able to hire workers at a lower wage than brown firms.

Worker  $i$ 's willingness to pay, let us denote it by  $\phi_i$ , can be defined implicitly as the wage difference that would make  $i$  indifferent between working in a brown or a green firm. The only variables in  $i$ 's utility function affected by  $i$ 's choice of firm type are his consumption (since wages may vary between firm types) and self-image. Since utility is linearly separable and effort is independent of firm type, as shown above, we can define  $\phi_i$  implicitly as

$$w(B) - \phi_i + \alpha_i V(\mathbf{x}(e_i, G), \mathbf{e}_i, E(G)) = w(B) + \alpha_i V(\mathbf{x}(e_i, B), \mathbf{e}_i, E(B))$$

or

$$\phi_i = \alpha_i [V(\mathbf{x}(e_i, G), \mathbf{e}_i, E(G)) - V(\mathbf{x}(e_i, B), \mathbf{e}_i, E(B))]. \quad (11)$$

Thus, a worker's willingness to pay will be positive only if he thinks social welfare would be higher if all firms were green than if all firms were brown; if abatement were socially wasteful, willingness

to pay for working in a green firm would be negative. In the latter case, green firms would never be able to survive in equilibrium: They would have to pay both the abatement cost and higher wages, and, in accordance with the screening argument provided below, their workers would in fact be less hard-working. Thus, the interesting case is when abatement is socially beneficial and the willingness to pay is, consequently, positive.

The environmental improvement if everyone worked in a green firm, compared to the case where everyone works in a brown firm, is  $Z$ . The social value of this is  $N\gamma Z$ . However, if everyone worked in green firms, wages per worker would have to be  $A/L$  lower, in order to cover the abatement cost. This means that the hypothetical welfare difference  $[V(\mathbf{x}(e_i, G), \mathbf{e}_i, E(G)) - V(\mathbf{x}(e_i, B), \mathbf{e}_i, E(B))]$  is independent of  $\alpha_i$ , implying, by (11), that whenever abatement is socially beneficial,  $\phi_i$  is increasing in the worker's moral motivation.

**Proposition 2** *Assume that  $\gamma Z > \frac{A}{L}$ , i.e. abatement is socially desirable. Then,  $i$ 's willingness to pay for working in a green firm,  $\phi_i$ , is a strictly increasing function of this individual's moral motivation  $\alpha_i$ :*

$$\phi_i = \phi(\alpha_i) = \alpha_i N \left[ \gamma Z - \frac{A}{L} \right]. \quad (12)$$

**Proof.** See the Appendix. ■

It is perhaps 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 wage costs. Our main conclusion, however, is actually much stronger than this. As we will demonstrate below, green firms may be capable not just of surviving, but possibly even of capturing the entire market, even if workers' willingness to pay equals zero for a substantial share of the work force. The crucial feature is that for an equal wage, some fraction of the workers would strictly prefer green firms; and these workers exert more effort than the average worker. Even with a quite marginal level of willingness to pay, this allows for labor market screening. Consequently, green firms may survive not primarily due to lower wages, but because they are able to attract more productive workers. Let us now turn to this issue.

## 6 Attracting productive workers: Market equilibrium

A worker  $i$  prefers working in a green firm if

$$w(G) + \phi(\alpha_i) \geq w(B). \quad (13)$$

Hence, if green firms pay a strictly lower wage than brown firms, only individuals with a relatively high moral motivation will apply for jobs in green firms. Since these individuals work harder, green firms will have a productivity advantage over the brown firms. That is, the firm's social responsibility profile can work as a screening device to attract responsible employees who shirk less than others.

Let  $\alpha$  be any threshold such that every  $i$  with  $\alpha_i \geq \alpha$  prefers to work in a green firm, while every  $i$  with  $\alpha_i < \alpha$  prefers a brown firm. For any distribution of  $\alpha_i$ , equation (10) implies that the average effort in each firm type  $\tau$  can be written as a function of the threshold  $\alpha$ ,  $e^\tau(\alpha)$ .<sup>17</sup> This gives us the following:

**Lemma 2** *Assume that there exist at least two individuals  $i$  and  $j$  such that  $\alpha_i \neq \alpha_j$ . Then, for every threshold  $\alpha \in [0, \bar{\alpha}]$ , average effort in green firms ( $e^G(\alpha)$ ) is strictly higher than average effort in brown firms ( $e^B(\alpha)$ ).*

**Proof.** See Appendix A. ■

Green firms, thus, have two advantages over brown firms: They have lower wage costs, and they have a higher productivity. They also, however, have the disadvantage of higher abatement costs.

The cost of being green,  $A$ , is always the same. The gains, however, will depend on the distribution of  $\alpha_i$ , and on  $\mu$ , the importance of unobservable effort for firm productivity. For the individual firm, the gains may also depend on how many other firms are green: If no green firms exist, and one firm turns green, it will attract the most highly motivated workers of all; if every firm is green, and one of them turns brown, it will attract only the very *least* motivated workers.

Define an equilibrium as a situation in which all existing firms have zero profit, and no worker wants to change his effort level or move to another type of firm. To study whether each type of firm

---

<sup>17</sup>Specifically,  $e^B(\alpha) = (c'(e_i))^{-1}(\alpha_i N \mu / (1 + \alpha_i N))$  (for all  $i$  such that  $\alpha_i \geq \alpha$ ), while  $e^G(\alpha) = (c'(e_i))^{-1}(\alpha_i N \mu / (1 + \alpha_i N))$  (for all  $i$  such that  $\alpha_i < \alpha$ ).

can exist in equilibrium, we must compare the compensation workers *demand* in order to accept brown employment,  $\phi(\alpha_i)$ , to the compensation brown firms are able to *offer*.

Hence, let  $\Delta w(\alpha)$  denote the *brown firms' compensating ability* as a function of any threshold  $\alpha$ . That is,  $\Delta w(\alpha)$  is the maximum wage differential  $w(B) - w(G)$  brown firms would be able to offer workers without making negative profits, in the event that every  $\alpha_i \geq \alpha$  worked for a green firm and every  $\alpha_i < \alpha$  worked for a brown firm.<sup>18</sup>

If both firm types exist in equilibrium, zero profits imply that wages must depend on average effort in the following way (see eq. 3):

$$w(G) = \frac{1}{L}[\mu(1 + e^G(\alpha)) - A] \quad (14)$$

$$w(B) = \frac{1}{L}\mu(1 + e^B(\alpha)) \quad (15)$$

Thus, brown firms' compensating ability is given by

$$\Delta w(\alpha) = \frac{1}{L}[A - \mu(e^G(\alpha) - e^B(\alpha))]. \quad (16)$$

where the shape of  $\Delta w(\alpha)$  depends on the specification of  $c(e_i)$ , on  $\mu$ , and on the distribution of  $\alpha_i$ .<sup>19</sup>

We are now prepared to state our main result.

**Proposition 3** *Assume that there exist individuals  $i$  and  $j$  such that  $i \neq j$  and  $\alpha_i \neq \alpha_j$ . Then there exists an  $\psi > 0$  such that whenever  $A < \psi\mu$ , there is a unique equilibrium with only green firms.*

**Proof.** See Appendix A. ■

Hence, even with extremely weak assumptions about the distribution of  $\alpha_i$ , and consequently about workers' willingness to pay, all firms will be green in equilibrium if only cleaning costs are sufficiently small, or if unobservable effort is sufficiently important for productivity.

---

<sup>18</sup>If green firms can offer higher wages than brown firms,  $\Delta w(\alpha)$  is negative, and all firms will be green. In the case where no brown firms exist, we define  $\Delta w(\alpha)$  as the maximum extra wage an entering 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  $\alpha = \bar{\alpha}$ .

<sup>19</sup>In the numerical example used in the figures below, the slope of  $\Delta w(\alpha)$  is negative, but it could also be positive: A very high  $\alpha$  means that there are only a few green firms, employing workers with unusually strong moral motivation; but at the same time, brown firms' productivity is relatively high too, because brown firms employ a large share of the workforce, including workers with a relatively high moral motivation.



Specifically, even if a substantial share of workers have no moral motivation whatsoever ( $\alpha_i = 0$ ), all firms can be green in equilibrium. To see this, imagine one brown entrant in a world of only green firms, offering the same wage as the green firms. This entrant can survive even if its productivity is slightly lower than green firms, since it does not have to pay the abatement cost; however, the entrant is able to attract only workers with  $\alpha_i = 0$ , that is, the *very least* motivated workers, so it may well end up with *much* lower productivity than other firms.<sup>20</sup>

If we assume that the population is sufficiently large to allow the distribution of  $\alpha_i$  to be approximated by a continuous probability distribution, the following Proposition specifies conditions under which there exists a unique equilibrium with, respectively, both firm types, only green firms, or only brown firms:

**Proposition 4** *Assume that the distribution of  $\alpha_i$  can be approximated by a continuous probability distribution assigning strictly positive probability to every  $\alpha_i \in [0, \bar{\alpha}]$ , and, moreover, that  $\partial\phi(\alpha)/\partial\alpha > \partial\Delta w(\alpha)/\partial\alpha$  for every  $\alpha \in [0, \bar{\alpha}]$ . Then there exists a unique labor market equilibrium characterized by an equilibrium threshold  $\alpha^*$ , such that every worker with  $\alpha_i > \alpha^*$  is employed by a green firm, every worker with  $\alpha_i < \alpha^*$  is employed by a brown firm, and where the following holds:*

$$\text{If } \alpha^* \in (0, 1) : \Delta w(\alpha) = \phi(\alpha) \tag{17}$$

$$\text{If } \alpha^* = 0 : \Delta w(0) \leq \phi(0) = 0$$

$$\text{If } \alpha^* = \bar{\alpha} : \Delta w(\bar{\alpha}) \geq \phi(\bar{\alpha}).$$

**Proof.** See Appendix A. ■

If the assumption  $\phi'(\alpha) > \Delta w'(\alpha)$  does not hold, multiple equilibria may occur, some of which may be unstable. This could happen for sufficiently convex effort costs  $c(e)$ . However, the main result still holds: a *unique* equilibrium with only green firms will always exist for sufficiently large  $\mu$  (or sufficiently small  $A$ ). For the interested reader, we have provided a more detailed discussion in Appendix B.

---

<sup>20</sup>Note that if there had been no labor market screening (e.g. because all workers were completely identical), then the existence of a non-negligible share of workers with no moral motivation would always imply the existence of at least one brown firm.

Figure 1 depicts workers' willingness to pay (the thick broken line) as a function of  $\alpha_i$ , whereas the solid line illustrates the maximum wage difference when  $\alpha = \alpha_i$ .<sup>21</sup> Here, all  $i$  with  $\alpha_i > \alpha^*$  will be employed by green firms (in this case, about half of the workers), while the rest are employed by brown firms.

Figure 1 about here

There are two factors influencing brown firms' compensating ability, i.e. their ability to offer a higher wage than green firms: First, green firms must pay the abatement cost (the first term in (16)). Second, expected average effort differs between firm types, due to labor market screening (the last term in (16)). Without screening, green firms might still be able to survive, but this may require a substantial willingness to pay by workers: the second term in (16) would disappear, and brown firms' compensating ability would be constant and equal to  $\frac{A}{L}$  (see Figure 1). In other words, without screening, survival of green firms would require that each worker in a green firm is prepared to pay his full share of the abatement cost. For the parameter values used in the figure, willingness to pay is not that large, implying that without screening, all firms would be brown; it is labor market screening which secures the survival of a substantial number of green firms.

If abatement is too costly, relatively to the potential gains of being green, all firms will be brown in equilibrium. Similarly, it can also be the case that all firms are green in equilibrium. This will happen if the importance of unobservable effort for firm productivity is sufficiently large.

The more important effort is for production, the larger is the importance of screening. In fact, a shift in  $\mu$ , which measures how important effort is for firm productivity, 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, except

---

<sup>21</sup>All figures below are based on the assumption that  $c(e) = e^\theta$  with  $\theta = 1.5$ , that  $\alpha_i$  is uniformly distributed, and  $\mu = 1.2$ . This secures a unique equilibrium. With very high  $\theta$ , multiple equilibria could arise.

$\mu$ , which is *lower* in Figure 2a and *higher* in Figure 2b.<sup>22</sup> The implication is that in the first situation (Figure 2a), there are only brown firms, while in the second case (Figure 2b), there are only green firms in equilibrium. Note that willingness to pay and abatement costs are the same in these two cases; only the importance of effort is different.

Consequently, if the importance of non-observable effort increases over time, one would expect social corporate responsibility to become more widespread. Similarly, one would expect to see a larger share of socially responsible firms in industries where production is crucially based on non-observable effort by employees.

Could this economy produce too much abatement? No: As noted above, willingness to pay for working in green firms is positive only if abatement is socially optimal. If  $\gamma Z < \frac{A}{L}$ , i.e. if the disutility of pollution is too small to justify the abatement cost, workers would in fact consider *brown* firms most socially beneficial, and there would be no green firms in equilibrium.

Now, if there exists an  $i$  with  $\alpha_i = \alpha^*$  such that  $\Delta w(\alpha^*) = \phi(\alpha^*)$ , then this will be a labor market equilibrium. In such an equilibrium, every worker with  $\alpha_i \geq \alpha^*$  will be employed by green firms, while any worker with  $\alpha_i < \alpha^*$  is employed by a brown firm. If  $\alpha^* \in (0, \bar{\alpha})$ , both firm types exist in equilibrium, and workers will self-select into green and brown firms, according to the strength of their moral motivation.

## 7 Policy analysis

Until now, we have assumed that there is no environmental policy. The government can, of course, use taxes, subsidies, or other instruments to stimulate abatement. However, if labor market screening provides a productivity advantage for green firms, less powerful policy instruments than otherwise will be needed to achieve a given environmental quality. In particular, the government can make all firms turn green simply by subsidizing abatement equipment, and, due to labor market screening, the required subsidy to achieve this is strictly lower than the abatement cost.<sup>23</sup>

<sup>22</sup>In Figure 2a,  $\mu = 0.8$ , while in Figure 2b,  $\mu = 1.6$ .

<sup>23</sup>Since emissions per firm are fixed in the present model, and abatement is a discrete decision for the firm, a subsidy on abatement equipment is formally equivalent to a marginal emission tax combined with a lump-sum transfer to each

Since highly motivated workers self-select into green firms, average productivity is always larger in green firms, provided that moral motivation differs between workers at all. This holds even when almost every firm is green, since brown firms will then only be able to recruit the very least motivated. Thus, a subsidy does not have to cover the entire abatement cost to drive brown firms out of business; it is sufficient that the subsidy covers the abatement cost less green firms' productivity advantage  $\mu(e^G(0) - e^B(0))$ . Moreover, provided that abatement is indeed socially desirable, the subsidy per firm required to make *every* firm green is strictly lower than the social value of the environmental disutility caused by each brown firm.

**Proposition 5** *Assume that the conditions for Propositions 2 and 4 hold. Then, a subsidy  $\Omega < A < \gamma ZL$  will be sufficient to make all firms green.*

**Proof.** See Appendix A. ■

There may also be another important role for policy in the current context. Above, we have implicitly assumed that workers have perfect knowledge about firms' social responsibility. However, information about such matters will in practice often be imperfect; and in that case, firms may have a strong incentive to pretend being green, but without paying the abatement cost  $A$ . If workers recognize this incentive, but are unable to distinguish truly responsible firms from cheaters, the screening mechanism described above could dissolve. Public disclosure of reliable information concerning firms' social responsibility may thus be crucial to allow labor market competition to favor green firms.<sup>24</sup>

---

firm.

<sup>24</sup>This may provide one possible explanation for the widespread practice among regulators to use seemingly lax sanctions towards violators, such as informal warnings or very low fines (Russell 1990, Nyborg and Telle 2004). As long as these sanctions are made public, they imply disclosure of information discrediting the firm's social responsibility. We have shown above that a subsidy on green firms, or a corresponding tax on brown firms, can potentially make all firms green even when the subsidy does not cover the abatement cost. Correspondingly, a relatively small expected sanction could be sufficient to deter violation of environmental regulations when the more productive workers are drawn to firms with a reputation for high social responsibility.

## 8 Variable emissions

In the above analysis, green firms have two advantages over brown firms, which may or may not outweigh green firms' abatement costs: Firstly, green firms are able to pay lower wages and still attract workers; secondly, the workers they recruit are more productive.

Above, we assumed that emissions per firm (before abatement) were fixed. If emissions had been increasing in the firm's production, however, green firms would have a third advantage: Any given worker with  $\alpha_i > 0$  would then provide strictly higher effort in a green than in a brown firm.<sup>25</sup>

When determining his effort level, the morally motivated worker asks himself: "What would happen if everybody exerted the same effort as me?" If he works hard, and everybody did so too, consumption for everyone would increase; this thought encourages him to work harder. However, if he works in a brown firm, such increased effort by everyone would also lead to a deterioration of environmental quality, and this would at least partially offset the former encouragement effect. In other words, if one's efforts contribute not only to increased productivity, but also to some negative externality, this undermines the moral worker's motivation to work hard.

Thus, assume now that pollution are given by  $zy^\tau$ , where  $z > 0$  denotes emission per unit of output and where  $y^\tau$  is production in a firm of type  $\tau$  as defined by eq. (2). Environmental quality is now given by

$$E = E^0 - bzy^B \quad (18)$$

where  $b$  is the share of brown firms, as above. It follows that

$$dV[\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i)]/de_i = \begin{cases} N[\mu - \mu L\gamma z - c'(e_i)] & \text{for } \tau = B \\ N[\mu - c'(e_i)] & \text{for } \tau = G \end{cases} \quad (19)$$

Inserting this into the first order condition (8) shows that a given individual  $i$  will exert different effort in green and brown firms:

$$\begin{aligned} e_i^G &= (c'(e_i))^{-1} \left( \frac{\alpha_i N \mu}{1 + \alpha_i N} \right) \\ e_i^B &= (c'(e_i))^{-1} \left( \frac{\alpha_i N \mu}{1 + \alpha_i N} (1 - L\gamma z) \right) \end{aligned} \quad (20)$$

---

<sup>25</sup>See Brekke and Nyborg (2004) for a full analysis.

Since  $L\gamma z > 0$ , this implies that for any given  $\alpha_i$ , effort is higher if the individual works in a green than in a brown firm.<sup>26</sup> Variable emissions would thus reinforce the result that green firms are more efficient than brown firms, making the survival of green firms more likely.<sup>27</sup>

One implication is that even if there were, for some reason, no labor market screening, such that the average  $\alpha_i$  were identical in brown and green firms, green firms would still have more productive workers: In a brown firm, the work motivation of any individual with  $\alpha_i > 0$  would be reduced, because working hard contributes to deteriorating the natural environment.

## 9 Concluding remarks

Our analysis has demonstrated that firms may be able to use their social corporate responsibility profile as a screening device to attract more productive workers. Consequently, green firms may be able to survive in the long run, even if workers' willingness to pay for green employment is rather limited. The screening mechanism could even be powerful enough to drive all brown firms out of business, even when a large number of workers have no moral motivation at all.

Several researchers have attempted to identify an empirical relationship between firms' environmental and economic performance, with somewhat mixed results (see, for example, Telle (2006) and the references therein). Here, we have assumed that profitable firms (whether green or brown) are imitated by others, until the extra earning potential has been exhausted, while unprofitable firms will vanish. Consequently, if both green and brown firms exist in equilibrium, our model provides no reason to expect their profitability to differ.

Our model would predict, however, that the share of green firms in equilibrium is decreasing in the

---

<sup>26</sup>Note that if marginal utility of income were decreasing in the income level, this result would be reinforced. The intuitive reason is that brown firms pay higher wages. When considering the self-image benefits of working harder, the individual weighs the incremental consumption benefits if everyone worked a little harder against the disutility of effort which everyone would also experience in that hypothetical case. If the utility of income were decreasing, the marginal utility of an hypothetically increased consumption would be lower in brown than in green firms.

<sup>27</sup>Willingness to pay would in fact be influenced as well: Workers comparing the benefits of working in each firm type would have to take into account that their actual effort would depend on firm type. This complicates the equilibrium analysis considerably (see Brekke and Nyborg 2004), but does not change the main insight.

abatement cost, but increasing in the importance of unobservable effort for firm productivity. Hence, if unobservable effort becomes more important over time, for example due to an increasing reliance on employees' highly specialized know-how, we would expect the share of green firms to increase over time. Similarly, in industries where unobservable effort is particularly important for firms' profitability, we would expect a relatively large share of socially responsible firms.

In our model, corporate social responsibility leads to labor market screening because some workers strictly prefer to work in a green rather than a brown firm, and because the strength of this preference is positively correlated with worker productivity. We have shown that this correlation arises endogenously when workers' self-image is based on a specific general principle of ethics. Similar correlations can arise from other ethical principles as well. For example, a preference to be *important to others* (Brekke and Nyborg 2006, Nyborg 2007), that is, a self-image that is increasing in the individual's own contribution to others' welfare, can induce a similar correlation. While a formal demonstration of this would require a separate analysis, the key is that an individual who cares a lot about contributing to social welfare typically derives a private benefit both from behaving environment-friendly and from contributing to her team. Moreover, these private benefits are smaller for individuals who generally care less about being important to others.

One should not, however, draw the conclusion that workers' moral motivation provides an easy and satisfactory solution to society's environmental and/or shirking problems. Although morally motivated workers partially internalize external effects, the internalization will be less than perfect, perhaps substantially less than perfect. Above, we demonstrated that effort levels will, in spite of the moral motivation, be suboptimal; in a model with continuous abatement, abatement expenditures would presumably also, in general, be suboptimal.<sup>28</sup>

Our intention is thus not to argue that environmental policy is redundant in the presence of workers' moral motivation. We have shown that the government can make every firm become green through a subsidy on abatement, and that the subsidy required to achieve this is strictly lower than the abatement cost, and also strictly lower than the social value of the environmental disutility caused

---

<sup>28</sup>Within a slightly different model, Brekke et al. (2003) demonstrated that there is undersupply of public goods even if moral motivation is very strong.

by each brown firm. Information disclosure may be another important task for the government, since limited verifiability of firms' social responsibility efforts could severely limit the potential of screening mechanisms like the one described here.

A possible extension of our model is to look further into the issue of fairness. Above, we assumed that pure rents to capital owners are zero in the long-run equilibrium, implying that increased average productivity benefits workers through higher wages. However, in some contexts it seems plausible that workers would think, instead, that if productivity increased, capital owners (or CEOs) would reap the gains for themselves, which would not necessarily be considered equally socially desirable as a wage increase for all. Thus, within the logic of our model, the distribution of firm profits could have profound implications for employees' work morale.

## References

- [1] Akerlof, G. A., and R. E. Kranton (2000): Economics and Identity, *Quarterly Journal of Economics* **115** (3), 715-753.
- [2] Alger, I., and C. A. Ma (2003): Moral Hazard, Insurance, and Some Collusion, *Journal of Economic Behavior and Organization* **50**, 225-247.
- [3] Altmann, S., T. Dohmen, and M. Wibral (2007): Do the Reciprocal Trust Less? Paper presented at the 62nd European Meeting of the Econometric Society, Budapest, August 2007.
- [4] Arora, S., and S. Gangopadhyay (1995): Toward a Theoretical Model of Voluntary Overcompliance, *Journal of Economic Behavior and Organization* **28**, 289-309.
- [5] Audi, R. (ed.) (1995): *The Cambridge Dictionary of Philosophy*, Cambridge, UK: Cambridge University Press.
- [6] Baron, D. (2005): Corporate Social Responsibility and Social Entrepreneurship, Research Paper No. 1916, Research Paper Series, Stanford Graduate School of Business.



- [7] Benabou, R., and J. Tirole (2006): Incentives and Prosocial Behavior, *American Economic Review* **96** (5), 1652-1678.
- [8] Besley, T., and M. Ghatak (2005): Competition and Incentives with Motivated Agents. *American Economic Review* 95(3), 616-636.
- [9] 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.
- [10] Blanco, M., D. Engelmann, and H.T. Normann (2007): A Within-Subject Analysis of Other-Regarding Preferences, Mimeo, Royal Holloway, University of London, [http://www.frisch.uio.no/behave\\_seminar/papers/Engelmann.pdf](http://www.frisch.uio.no/behave_seminar/papers/Engelmann.pdf).
- [11] Brekke, K. A., S. Kverndokk, and K. Nyborg (2003): An Economic Model of Moral Motivation, *Journal of Public Economics* 87 (9-10), 1967-1983.
- [12] Brekke, K.A., and K. Nyborg (2004): Moral Hazard and Moral Motivation: Corporate Social Responsibility as Labor Market Screening, Memorandum 25/2004, Department of Economics, University of Oslo.
- [13] Brekke, K.A., and K. Nyborg (2006): Selfish Bakers, Caring Nurses? A Model of Work Motivation. Paper presented at the HERO/HEB Workshop on Health Economics, Oslo, August-September 2006.
- [14] Bruvoll, A., B. Halvorsen, and K. Nyborg (2002): Households' Recycling Efforts, Resources, Conservation & Recycling 36 (4), 337-354.
- [15] Cullis, J., Lewis, A., Winnett, A., 1992. Paying to be good? UK ethical investments. *Kyklos* 45, 3-24.
- [16] Frank, Robert H. (2003): *What Price the Moral High Ground? Ethical Dilemmas in Competitive Environments*, Princeton University Press.

- [17] Heal, G. (2005): Corporate Social Responsibility: An Economic and Financial Framework. *The Geneva Papers on Risk and Insurance* 30(3), 387-409.
- [18] Heyes, A. (2005a): A Signaling Motive for Self-Regulation in the Shadow of Coercion, *Journal of Economics and Business* 57, 238-246.
- [19] Heyes, A. (2005b): The Economics of Vocation or ‘Why is a Badly Paid Nurse a Good Nurse’? *Journal of Health Economics* 24 (3), 561-569.
- [20] Holmstrom, B. (1982): Moral Hazard in Teams, *Bell Journal of Economics* 13, 324-340.
- [21] 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.
- [22] Nyborg, K. (2007): I Don’t Want to Hear About it: Rational Ignorance among Duty-Oriented Consumers. Paper presented at the 15th EAERE Conference, Thessaloniki, June 2007.
- [23] Nyborg, K., and K. Telle (2004): The Role of Warnings in Regulation: Keeping Control with Less Punishment, *Journal of Public Economics* 88 (12), 2801-2816.
- [24] Reinikka, R., and J. Svensson (2003): Working for God? Evaluating Service Delivery of Religious Not-for-Profit Health Care Providers in Uganda. Policy Research Working Paper Series 3058, Washington, DC: The World Bank.
- [25] Russell, C. (1990): Monitoring and Enforcement, in P. Portney (ed.), *Public Policies for Environmental Protection*, Resources for the Future, Washington DC.
- [26] Stiglitz, J. E. (1975): The Theory of "Screening", Education, and the Distribution of Income, *American Economic Review* 65 (3), 283-300.
- [27] Telle, K. (2006): “It Pays to be Green” – A Premature Conclusion? *Environmental and Resource Economics* 35 (3), 195-220.
- [28] Vitell, S.J. and D.L. Davis (2004): The Relationship Between Ethics and Job Satisfaction: An Empirical Investigation, *Journal of Business Ethics* 9 (6), 489-494.

# A Proofs

## Proof of Lemma 1

**Proof.** Consider first the case where social welfare is given by  $V^S = \sum_{j=1}^N [u(x_j) + \gamma E - c(e_j) + S_j]$ .

Using (7), we can write this as  $V^S = \sum_{j=1}^N [u(x_j) + \gamma E - c(e_j) + \alpha_j V^S(e_i, \tau_i)]$ , where  $V^S(e_i, \tau_i)$  denotes  $V^S$  in the case that  $e_i = e_j$  for all  $j$  and  $\tau_i = \tau_j$  for all  $j$ . This implies that

$$V^S(e_i, \tau_i) = \sum_{j=1}^N [u(x_j(e_i, \tau_i)) + \gamma E(\tau_i) - c(e_i) + \alpha_j V^S(e_i, \tau_i)]$$

Rearranging, we get

$$\left(1 - \sum_{j=1}^N \alpha_j\right) V^S(e_i, \tau_i) = \sum_{j=1}^N [u(x_j(e_i, \tau_i)) + \gamma E(\tau_i) - c(e_i)]$$

implying that  $kV^S(e_i, \tau_i) = V(\mathbf{x}(e_i, \tau_i), \mathbf{e}_i, E(\tau_i))$ , where the constant  $k = \left(1 - \sum_{j=1}^N \alpha_j\right) > 0$  (since, by assumption,  $\sum_{j=1}^N \alpha_j < 1$ , which rules out extreme altruism). Thus, hypothetical social welfare including self-image, if everybody acted like  $i$ , is just a rescaling of hypothetical social welfare excluding self-image, if everybody acted like  $i$ . Furthermore, since hypothetical social welfare enters the utility function only through self-image, replacing  $V$  by  $V^S$  is formally equivalent to a rescaling of the moral motivation parameter  $\alpha_i$ . These two cases would thus be behaviorally indistinguishable. ■

## Proof of Proposition 2:

**Proof.** If both firm types exist, (5) and (3) imply

$$\begin{aligned} x(e_i, G) &= (1 + e^G)\mu - \frac{A}{L} \\ x(e_i, B) &= (1 + e^B)\mu \end{aligned} \tag{21}$$

where  $x(e_i, \tau_i)$  is individual income in the case that for every  $j$ ,  $e_j = e_i$  and  $\tau_j = \tau_i$ . Combining this with (6) and (11) gives

$$\phi_i = \alpha_i (N[x(e_i, G) + \gamma E(G)] - N[x(e_i, B) + \gamma E(B)]) \tag{22}$$

Combining this with (21) yields

$$\phi_i = \alpha_i N \left[ \left( \mu(1 + e_i) - \frac{A}{L} \right) + \gamma E(G) - \left( \mu(1 + e_i) \right) - \gamma E(B) \right] \tag{23}$$

$$= \alpha_i N \left[ -\frac{A}{L} + \gamma E(G) - \gamma E(B) \right] \tag{24}$$

Further, by (4),  $E(B) = E^0 - Z$ , while  $E(G) = E^0$ . Consequently,  $\phi_i = \alpha_i N[\gamma Z - \frac{A}{L}]$ . ■

**Proof of Lemma 2:**

**Proof.** Since there exist different individuals  $i \neq j$  with  $\alpha_i \neq \alpha_j$ , and since effort is increasing in  $\alpha_i$  (Proposition 1), it follows that  $e_i \neq e_j$ ; hence there exists variation in effort between workers. Recall that  $\alpha$  is defined as any threshold such that every  $i$  with  $\alpha_i \geq \alpha$  prefers to work in a green firm, while every  $i$  with  $\alpha_i < \alpha$  prefers a brown firm. Thus, green firms will hire those workers who exert the highest effort, while brown firms hire those who exert the least effort. It follows that  $e^G(\alpha) > e^B(\alpha)$  for any  $\alpha$ . ■

**Proof of Proposition 3:**

**Proof.** From Lemma 2, we know that when there exist individuals  $i \neq j$  such that  $\alpha_i \neq \alpha_j$ , then  $e^G(\alpha) > e^B(\alpha)$  for any  $\alpha$ . Thus we can define

$$\min_{\alpha} (e^G(\alpha) - e^B(\alpha)) = \psi > 0.$$

Now for any  $\alpha$ , it follows from (16)

$$\begin{aligned} \Delta w(\alpha) &= \frac{1}{L}[A - \mu(e^G(\alpha) - e^B(\alpha))] \\ &\leq \frac{1}{L}[A - \mu\psi] \leq 0 \end{aligned}$$

for  $\mu \geq A/\psi$ . ■

**Proof of Proposition 4:**

**Proof.** Under the conditions of the theorem,  $\phi(\alpha)$  and  $\Delta w(\alpha)$  are continuous functions with  $\phi'(\alpha) > \Delta w'(\alpha)$ .  $\phi(\alpha) - \Delta w(\alpha)$  is thus a strictly increasing function. It follows that either  $\phi(\alpha) > \Delta w(\alpha)$  for all  $\alpha$ ,  $\phi(\alpha) < \Delta w(\alpha)$  for all  $\alpha$ , or there is a unique  $\alpha^*$  such that  $\phi(\alpha^*) = \Delta w(\alpha^*)$ . In the first case, we have, in particular  $\phi(0) > \Delta w(0)$ . We know by eq. (12) that  $\phi(0) = 0$ ; thus, if all firms were green, a potential brown entrant would be able to pay less than the green firms, and all workers would strictly prefer the green firm, which is an equilibrium. In a similar fashion, it follows that all firms being brown is an equilibrium when  $\phi(\alpha) < \Delta w(\alpha)$  for all  $\alpha$ . If there exists an  $\alpha^* \in [0, \bar{\alpha}]$  such that  $\phi(\alpha^*) = \Delta w(\alpha^*)$ , then at this wage difference all individuals with  $\alpha_i > \alpha^*$  have a willingness to pay  $\phi(\alpha_i) > \Delta w(\alpha^*)$  and hence strictly prefer working in a green firm, every  $i$  with  $\alpha_i < \alpha^*$  strictly prefers

a brown firm, and every  $i$  with  $\alpha_i = \alpha^*$  is indifferent. This proves the existence of an equilibrium.

To prove uniqueness, note that an  $\alpha$  with  $\phi(\alpha) < \Delta w(\alpha)$  cannot constitute an equilibrium if there are workers  $i$  with  $\alpha_i > \alpha$ : If  $\alpha$  is an equilibrium value of the threshold, this means (by definition) that all workers with  $\alpha_i > \alpha$  prefers working in a green firm; while we know that every  $i$  such that  $\phi(\alpha_i) < \Delta w(\alpha)$  prefers working in a brown firm. Similarly, any  $\alpha$  with  $\phi(\alpha) > \Delta w(\alpha)$  cannot be an equilibrium if there are workers with  $\alpha_i \leq \alpha$ , since these workers will not prefer the brown firms.

■

### Proof of Proposition 5:

**Proof.** In the case with no subsidy, we know from eq.(16) that  $\Delta w(\alpha) = \frac{1}{L}[A - \mu(e^G(\alpha) - e^B(\alpha))]$ . A subsidy  $\Omega$  on the purchase of abatement equipment is, from firms' perspective, equivalent to a lower purchase price. Hence, in the case with a subsidy, brown firms' compensating ability is strictly negative if  $A - \Omega < \mu(e^G(\alpha) - e^B(\alpha))$ . We are considering the minimum subsidy required to drive *all* brown firms out of business. Denote this subsidy level  $\Omega^G$ . If brown firms are driven out of business, we must have  $\alpha^* = 0$ ; moreover, brown firms' compensating ability in this situation must be negative. This holds whenever  $A - \Omega^G < \mu(e^G(0) - e^B(0))$ . Due to Lemma 2, we know that  $e^G(\alpha) - e^B(\alpha) > 0$ ; hence  $e^G(0) - e^B(0) > 0$ . Thus, all firms will be green at a subsidy level  $\Omega^G > A - \mu(e^G(0) - e^B(0)) < A$ .

Furthermore, from Proposition 2, we know by assumption that  $\gamma Z > \frac{A}{L}$ . It follows that  $\gamma ZL > A$ , and, consequently, that  $\Omega^G < A < \gamma ZL$ . To see that  $\gamma ZL$  is the value of the environmental damage caused by each brown firm, note first that the number of firms is  $N/L$ . The environmental damage caused by each brown firm is, by (4),  $Z\frac{L}{N}$ . The value of this for a single individual is  $\gamma Z\frac{L}{N}$ , and, consequently, its value for the entire society of  $N$  individuals is  $\gamma ZL$ . ■

## B Multiple equilibria

If the assumption  $\phi'(\alpha) > \Delta w'(\alpha)$  does not hold, multiple equilibria may occur, some of which may be unstable. This could happen for sufficiently convex costs of effort  $c(e)$ . However, the general insight is that a *unique* equilibrium with only green firms will always exist for sufficiently large  $\mu$ .

Define an equilibrium as a situation in which all (existing) firms have zero profit, and no worker wants to move to another type of firm. Any equilibrium must be characterized by a threshold value  $\alpha$  such that all  $i$  with  $\alpha_i \geq \alpha$  prefer green and the rest prefer brown. (For  $\alpha = \bar{\alpha}$ , we assume that  $i$  with  $\alpha_i = \alpha$  prefer brown.) A threshold  $\alpha \in (0, 1)$  defines an equilibrium iff  $\Delta w(\alpha) = \phi(\alpha)$ . When  $\alpha = 0$ , there are no workers with  $\alpha_i < \alpha$ , and with  $\alpha = \bar{\alpha}$  there are none with  $\alpha_i > \alpha$ . Thus in these cases the equilibrium condition is one-sided. The general equilibrium conditions are as stated in Proposition 4, eq. (17).

An equilibrium is unstable if a slight deviation will induce a larger deviation. If  $\alpha$  is a stable equilibrium, then if the threshold for some reason moves to  $\alpha' < \alpha$ , then workers must have an incentive to move to brown firms, which requires that  $\Delta w(\alpha') > \phi(\alpha')$ . Thus, for a stable equilibrium  $\alpha$ , if  $\Delta w(\alpha) = \phi(\alpha)$  then  $\phi(\alpha') - \Delta w(\alpha')$  must be increasing in  $\alpha'$ .

Now if  $\Delta w(\alpha) - \phi(\alpha)$  is declining, there can be only one equilibrium. And as  $\phi(\cdot)$  is increasing, this will be the case unless  $\Delta w(\cdot)$  is sufficiently increasing. It turns out that  $\Delta w(\cdot)$  may be rapidly increasing when  $c(e)$  is strongly convex. In this case the cost of effort is low for small values of  $e$  but then rapidly increases. Most workers, except those with  $\alpha_i \approx 0$ , will then exert about the same effort. In the entire population there will only be a few workers with substantially lower effort. These workers will never constitute a large share of the labor force in green firms; hence average effort in green firms is not very sensitive to changes in  $\alpha$ . For brown firms, however,  $\alpha$  matters a lot. With  $\alpha \approx 0$ , brown firms will only hire among those with low effort, while for high  $\alpha$ , these low effort workers will constitute a small share of brown firms' employees. Hence, average effort in brown firms is rapidly increasing for small  $\alpha$ . As a consequence, the wage difference would be rapidly increasing for small  $\alpha$ , but would then flatten out. This is illustrated in Figure 3, where we have assumed that  $c(e) = e^\theta$  as in Figures 1-2, but in Figure 3,  $\theta = 3$ .

The case in Figure 3 exhibits three equilibria, one for  $\alpha = 0$  (all firms are green), one for  $\alpha \approx 0.1$ , and one with  $\alpha \approx 0.5$ . The equilibrium at  $\alpha \approx 0.1$  is unstable: If workers with a slightly higher value of  $\alpha_i$  moved to brown firms, brown firms' compensating ability would increase, they would attract even more workers, and this process would continue until  $\alpha \approx 0.5$ . If  $\mu$  increases,  $\Delta w(\alpha)$  will shift

down and eventually  $\Delta w(\alpha) - \phi(\alpha) < 0$  for all  $\alpha$ , and the only equilibrium will be that all firms are green, as stated in Proposition 3.

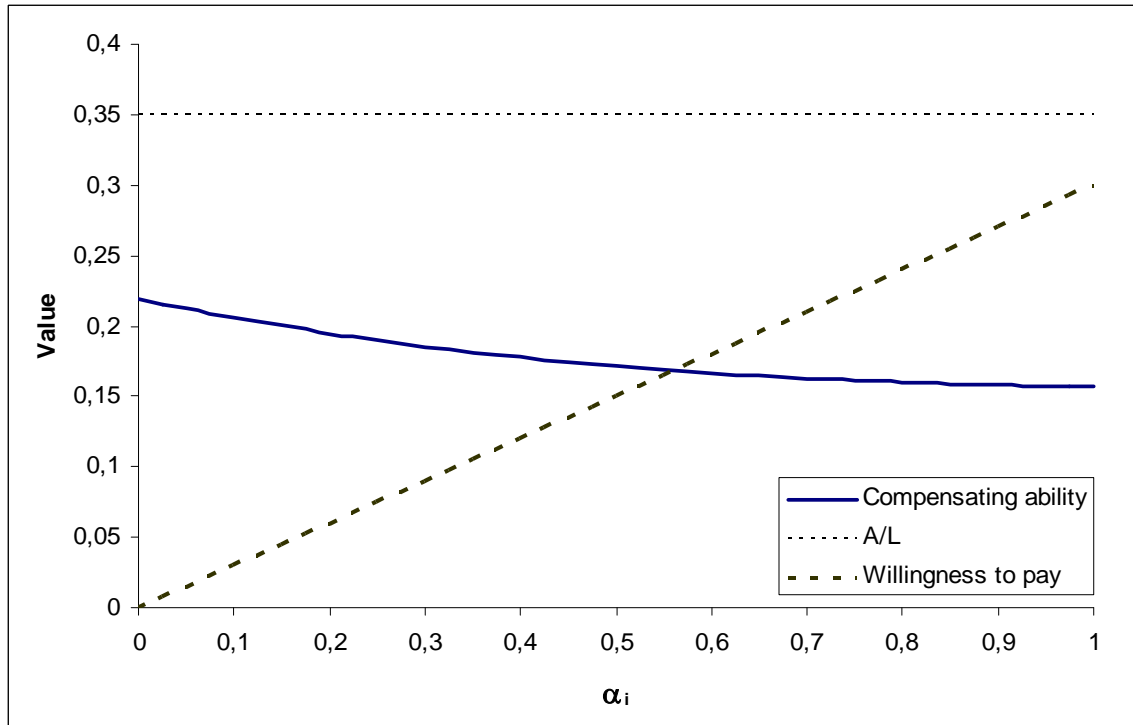


Figure 1: Brown firms' compensating ability,  $\Delta w(\alpha)$ , given  $\alpha=\alpha_i$ , and workers' willingness to pay,  $\phi(\alpha_i)$ .



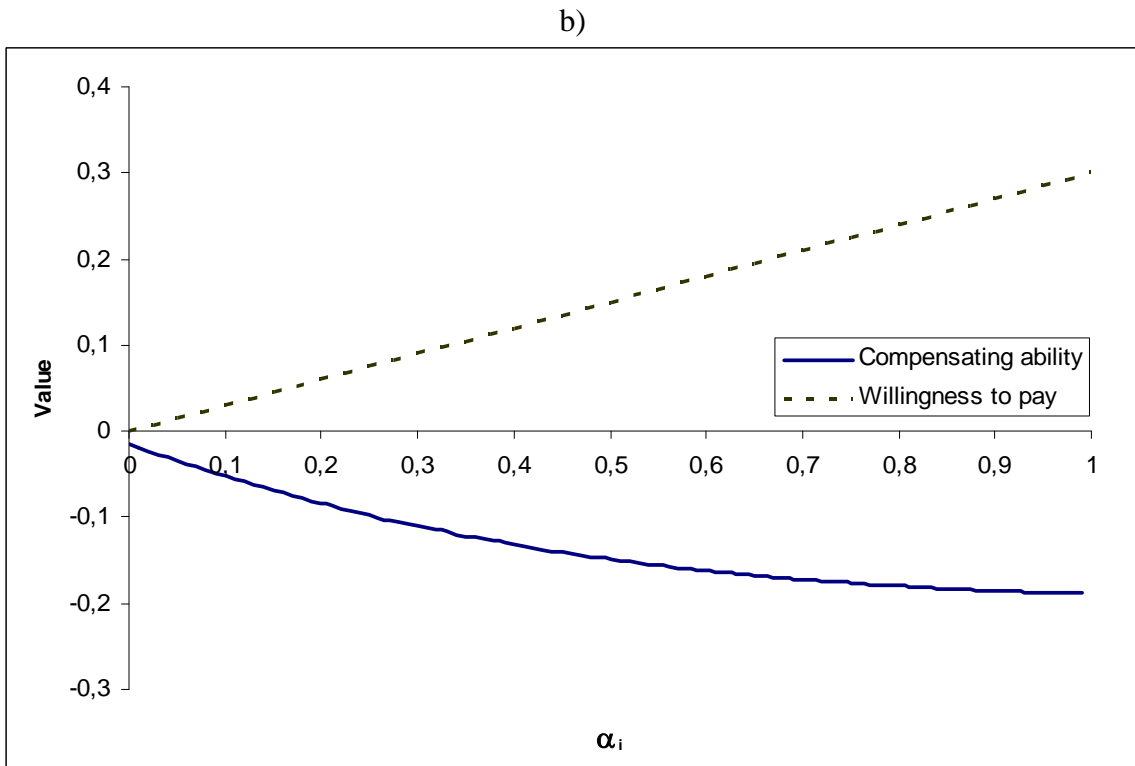
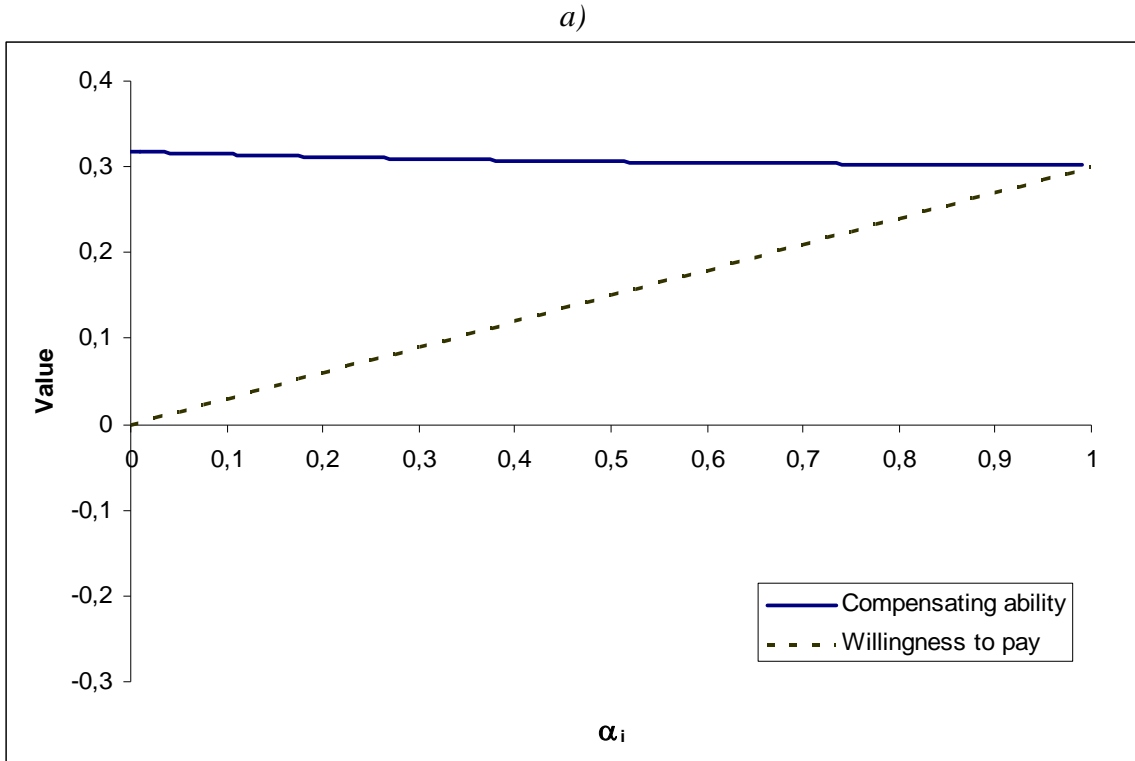


Figure 2: Brown firms' compensating ability,  $\Delta w(\alpha)$ , given  $\alpha = \alpha_i$ , and workers' willingness to pay,  $\phi(\alpha_i)$ , with varying importance of unobservable effort  $\mu$ . In (a),  $\mu = 0.8$ , and all firms are brown; in (b),  $\mu = 1.6$ , and all firms are green.

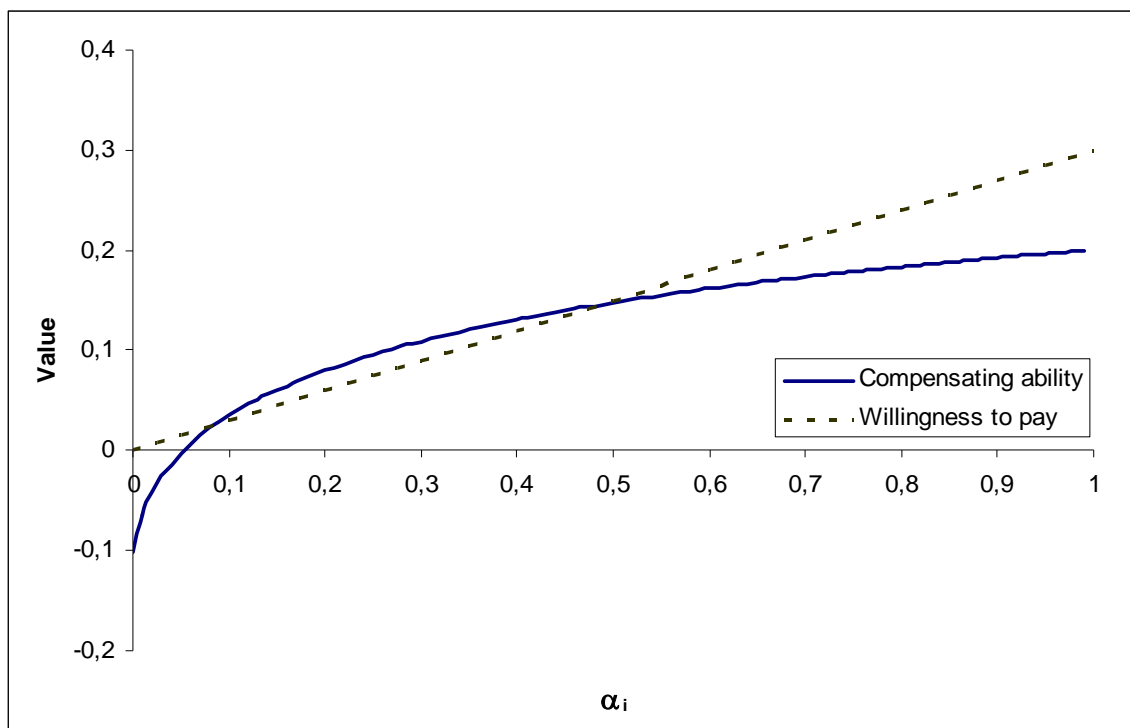


Figure 3: Multiple equilibria