Impact Trickles Down: Exit, Engagement and Firm-Stakeholder Relationships *

Briana Chang[†] and Harrison Hong[‡]

October 2025

Abstract

We develop a general equilibrium model to address the debate on whether exit or engagement by values-driven stakeholders is more effective in mitigating firm harm. Firm and stakeholders form relationships, trading off production complementarities and mitigation costs. Using a novel characterization of the matching equilibrium, we show how a values shock leads stakeholders to either optimally break (exit) or maintain prior relationships (engage). Treated high-productivity stakeholders choose exit, reallocating toward firms that already mitigate, whereas low-productivity stakeholder choose engagement. While exit has limited direct effect compared to engagement, it induces equilibrium reallocation by lower-productivity stakeholders unaffected by the shock, which in turn incentivizes new firms to mitigate. A calibration suggests that the aggregate impact of exit has been understated relative to engagement, once spillovers are accounted for.

*We thank Marcus Opp, Jan Starmans, Lasse Pedersen, David Lando, Navin Karthik, Qingmin Liu, Yeon-Koo Che, Lily Fang, Alminas Zaldokas, Wenxi Jiang, Lou Dong, Kai Li, Adam Zhang (discussant), Deeksha Gupta (discussant), Morris Davis, Kerry Back, Bruce Carlin, Gustavo Grullon, Stephanie Johnson, Kunal Sachdeva, Yuhang Xing, and seminar participants at the Stockholm School of Economics, Copenhagen Business School, INSEAD, Columbia University, Rice University, Peking University HSBC Business School Inaugural Finance Colloquium, Minnesota Corporate Finance Conference, the Baruch Sustainable Finance Conference, American Finance Association Meetings, Rutgers Business School, University of Zurich, FIRN Tasmania Conference and University of New South Wales for helpful comments.

[†]Hong Kong University of Science and Technology

[‡]Columbia University

1 Introduction

Employee walkouts at technology firms, shareholder climate resolutions at global banks, and boycotts of controversial brands illustrate how values-driven stakeholders increasingly pressure firms to mitigate environmental and social harm. These strategies typically take two forms: exit—divestment, boycotts, or quitting—and engagement, including employee and shareholder proposals and dialogue. Which strategy is more effective in reducing harm? This contemporary debate echoes Hirschman 1972's classic distinction between exit and voice as responses to organizational decline.

Yet despite decades of research, the aggregate impact of exit relative to engagement remains less understood, particularly when stakeholders and firms interact in general equilibrium. Recent influential work, such as Broccardo, Hart, Zingales, et al. 2022 and Dimson, Karakaş, and Li 2015, argue that engagement dominates exit when it comes to mitigating harm. Yet, large stakeholders, be they big banks like JP Morgan or large funds like Norges Bank Investment Management choose exit rather than engagement.[1] The impact of these divestments are controversial since these large stakeholders often form new relationships with firms that are already green.

We address this gap by developing a general equilibrium model of stakeholder–firm interactions. Firms and multiple groups of stakeholders (e.g., workers, banks, suppliers) are matched in competitive multi-sided markets, where relationships determine both output and the mitigation of environmental or social harm. Following a shock to the proportion of values-driven stakeholders in the economy, exit and engagement emerge as the optimal strategies for breaking or maintaining previous relationships. Our model therefore captures not only the direct effects of exit and engagement but also the reallocation spillovers that amplify their influence in equilibrium.[2]

[1]Examples include the famous 2015 Norges Bank divestment from coal companies or a wave of big banks during a similar period in time committing to divest from carbon intensive companies.

[2]Spillovers appear in a number of areas in economics including agglomeration effects of cities, (Ellison and Glaeser 1997), spillovers of firm entry (Greenstone, Hornbeck, and Moretti 2010), the spread of technological shocks through production networks (Acemoglu, Akcigit, and Kerr 2016), and macro-financial shocks (Huber 2023), to name a few. These studies share our focus on how small local changes can have aggregate consequences,

Methodologically, we contribute to the literature on multi-sided matching with multidimensional characteristics. While the existence and efficiency of such equilibria are well studied in general frameworks (e.g., Hatfield and Kominers 2015), full characterization is typically intractable when stakeholders differ across multiple dimensions. We introduce a novel iterative solution procedure that breaks down the team-formation problem into multiple rounds of bilateral matching. This approach allows us to solve for multi-sided equilibrium in a tractable way and to highlight how production complementarities, mitigation preferences, and harm that increases with output jointly shape stakeholder allocation.

This framework delivers several insights. First, because mitigation is non-rival for stakeholders on the same team, there are economies of scale from having more values-driven stakeholders within a firm. This induces values-driven stakeholders of comparable productivity to sort together and with less productive firms than they otherwise would in the absence of values. However, heterogeneity in values across stakeholder groups implies that some teams must mix values-driven and pecuniary stakeholders. Large firms, facing high mitigation costs from output, are especially incentivized to match with pecuniary stakeholders, while smaller firms are more likely to attract values-driven ones. Consequently, in equilibrium, values-driven stakeholders sort into less productive firms and pecuniary stakeholders into more productive ones. Above a critical productivity cutoff, firms are segmented, hiring either exclusively values-driven stakeholders (and fully mitigating) or exclusively pecuniary stakeholders (and not mitigating). Below this cutoff, pecuniary firms strategically adopt mitigation to compete for talented values-driven stakeholders, generating mixed teams.

We then consider the effects of a values shock, motivated by empirical differences between banks and workers. Historically, workers are more likely than banks to be values-driven. An increase in the fraction of values-driven banks shifts the equilibrium cutoff, making it easier for teams to consist solely of values-driven or pecuniary members. The key mechanism we uncover is that highly productive stakeholders optimally exit, reallocating to firms that already

though our contribution is to show that values-driven preference shocks operate through novel channels in stakeholder–firm matching.

mitigate. This exit has limited direct impact compared to engagement but triggers spillovers: less-productive stakeholders are also displaced and reallocate, inducing their new firms to adopt mitigation. By contrast, the least productive stakeholders who receive a values shock optimally engage, but without reallocation their impact is limited to the firm they remain with.

Our central finding is that the aggregate impact of exit has been substantially understated relative to engagement once spillovers are accounted for. While engagement delivers firm-level change, exit generates equilibrium adjustments that cascade across firms. A calibration using data on firms, banks, and workers demonstrates that these spillover effects are quantitatively significant.

The paper proceeds as follows. Section 2 situates our contribution within the literature on exit, engagement, and stakeholder influence. Section 3 presents the model. Section 4 presents properties of optimal teams. Section 5 characterizes the equilibrium. Section 6 considers comparative effects of exit and engagement following a values shock. Section 7 calibrates the model using data on banks and workers. Section 8 concludes.

2 Related Literature and Contribution

Our paper speaks to three main literatures: the economics of exit and voice, the role of values-driven preferences in markets, and the theory of matching and sorting.

Exit, voice, and organizational discipline. Hirschman 1972 established the canonical framework for how dissatisfied stakeholders discipline organizations. Economists have formalized exit as a disciplinary device in markets. Tiebout 1956 showed how mobility ("exit") disciplines local governments, and subsequent work in political economy studied how migration and emigration function as exit mechanisms (e.g., Epple, Romer, and Sieg 2001). In contrast, voice—through collective action or activism—has been harder to model formally, and thus exit has often dominated in economics. Our paper revisits this debate in a general equilibrium framework with multi-sided matching, highlighting that the aggregate effects of exit have been understated once spillovers are taken into account.

Exit versus engagement in markets and finance. Exit and engagement have been contrasted most directly in models of socially responsible investment. Socially responsible funds (exit) can affect firm financing (see, e.g., Heinkel, Kraus, and Zechner 2001, Hong, Wang, and Yang 2021, Pástor, Stambaugh, and Taylor 2021, Pedersen, Fitzgibbons, and Pomorski 2021, Oehmke and Opp 2023), while Broccardo, Hart, Zingales, et al. 2022 show that exit is a less efficient means to achieve impact than voting or voice. Our contribution is to provide a general equilibrium mechanism where exit, usually seen as blunt, produces large spillovers through reallocation that amplify its aggregate effect, challenging the consensus that engagement dominates.

Values, preferences, and non-pecuniary motives. A second strand of literature emphasizes that stakeholders may act on non-pecuniary motives. Besley and Ghatak 2005 formalize mission preferences in organizations, showing how workers with prosocial preferences sort into firms with aligned missions. Ellingsen and Johannesson 2008 and Bénabou and Tirole 2006 study prosocial motivation, signaling, and identity in economics, while Akerlof and Kranton 2000 integrate identity directly into utility. These works provide the microfoundations for values-driven behavior. We extend this literature by embedding values-driven stakeholders in a competitive general equilibrium with multi-sided markets, allowing us to evaluate how their exit versus engagement strategies affect aggregate outcomes.

Matching and sorting in general equilibrium. Methodologically, our work connects to the literature on matching with multidimensional heterogeneity. Unlike one-dimensional matching models (Sattinger 1979, Tervio 2008, Gabaix and Landier 2008), where the supermodularity of the surplus function simply governs the sorting condition — the characterization conditions are generally more complex in the multidimensional content. Several studies consider frameworks where multiple characteristics can be summarized by a single index so that the matching so that the matching is de facto one-dimensional. In the bilateral matching setting, previous work (Dupuy and Galichon 2014, Lindenlaub 2017, Chiappori, McCann, and Pass 2016, Chiappori, Oreffice, and Quintana-Domeque 2018) have provided full characterization under some specific properties under two-dimensional matching. We contribute a tractable iterative solution to multi-sided matching with values-driven preferences, allowing us to characterize equilibria that would otherwise be intractable.

In summary, our contribution is threefold. First, we provide a tractable model of multi-sided matching in general equilibrium with values-driven stakeholders, extending the canonical matching literature to incorporate non-pecuniary preferences. Second, we identify a new mechanism—spillovers from exit—that amplifies the aggregate impact of exit relative to engagement. This challenges the dominant view, grounded in both theory and practice, that engagement is the more effective channel. Third, we calibrate the model using data on banks and workers, showing that spillovers from exit are quantitatively significant.

3 Model

Production and harm. There are N types of stakeholders and one firm, indexed by $\ell \in L \equiv \{1, 2, \ldots, N+1\}$. Each stakeholder type $\ell \leq N$ has skill $x_\ell \in X_\ell$, while the firm has productivity $x_{N+1} \in X_{N+1}$. All types have unit mass, with smoothly distributed skills on compact supports.

Output depends multiplicatively on all agents' characteristics:

$$y(\mathbf{x}) = \prod_{\ell=1}^{N+1} x_\ell,$$

where $\mathbf{x} = (x_1, \ldots, x_{N+1})$. Production generates environmental or social harm $\sigma y(\mathbf{x})$, where $\sigma > 0$ is the harm rate. Firms may reduce harm by $m \geq 0$ at linear cost cm.

Stakeholder preferences. Each stakeholder matches to exactly one firm, capturing the notion of a bilateral relationship (e.g. a bank lending to one firm, or a worker employed at one firm).

Stakeholders differ in whether they are values-driven. Let $\theta_\ell \in \{0,1\}$ denote type. Utility is

$$u(p, e \mid \theta_\ell) = p - \theta_\ell \psi(e),$$

where p is the transfer received, e is firm harm, and $\psi(\cdot)$ is increasing and convex.

- Pecuniary stakeholders ($\theta_\ell = 0$) care only about compensation.
- Values-driven stakeholders ($\theta_\ell = 1$) also dislike the harm produced by their firm.

Thus stakeholder types are $a_\ell = (x_\ell, \theta_\ell)$, distributed with measure μ_ℓ on $X_\ell \times \{0,1\}$. The share of values-driven stakeholders of type ℓ is denoted by λ_ℓ.

Firms themselves are profit-maximizing and do not have non-pecuniary preferences, i.e. $a_{N+1} = (x_{N+1}, 0)$ and $\lambda_{N+1} = 0$.

Mitigation and surplus. For any team $\mathbf{a} = (a_1, \ldots, a_{N+1})$, let

$$n(\boldsymbol{\theta}) = \sum_{\ell=1}^{N+1} \theta_\ell$$

denote the *stakeholder-values index* of the firm, i.e. the number of values-driven members in the team.

The joint surplus is

$$\Lambda(\mathbf{a}) = \max_{m \geq 0} \; y(\mathbf{x}) - cm - n(\boldsymbol{\theta})\psi(\sigma y(\mathbf{x}) - m). \tag{1}$$

Intuitively, mitigation is chosen to maximize total surplus. If $n(\boldsymbol{\theta}) = 0$, the firm is purely pecuniary and does not mitigate. If $n(\boldsymbol{\theta}) > 0$, mitigation is positive, with costs shared across stakeholders via transfers.

Stakeholder payoffs. For any stakeholder a_ℓ, let the utility be

$$U_\ell(a_\ell) = \max_{\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}} \Lambda(\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}, a_\ell) - \sum_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'}). \tag{2}$$

That is, the payoff to stakeholder a_ℓ is the team surplus minus the equilibrium utilities of all other members. The firm's problem is a special case of (2) with $\ell = N + 1$.

Competitive equilibrium. A competitive equilibrium consists of:

1. A matching γ between firms and stakeholders, with marginal distributions equal to μ_ℓ for all ℓ (market clearing).
2. A mitigation policy $m^*(x_{N+1})$ for each firm.
3. Equilibrium utilities $\{U_\ell(a_\ell)\}$ for each stakeholder type.

These must satisfy:

- Each team $\mathbf{a}$ formed under γ maximizes joint surplus (1).
- Each stakeholder a_ℓ receives utility consistent with (2).
- The matching γ clears the market.

As shown in Hatfield and Kominers (2015), such competitive equilibria exist and correspond to stable outcomes in multilateral matching with transfers. Figure 1 summarizes the flow of our model.

4 Properties of Optimal Teams

We now analyze the surplus function and the resulting matching outcomes. Throughout, recall that a team is characterized by two sufficient statistics: its productivity $y(\mathbf{x})$ and its stakeholder-values index $n(\boldsymbol{\theta})$.

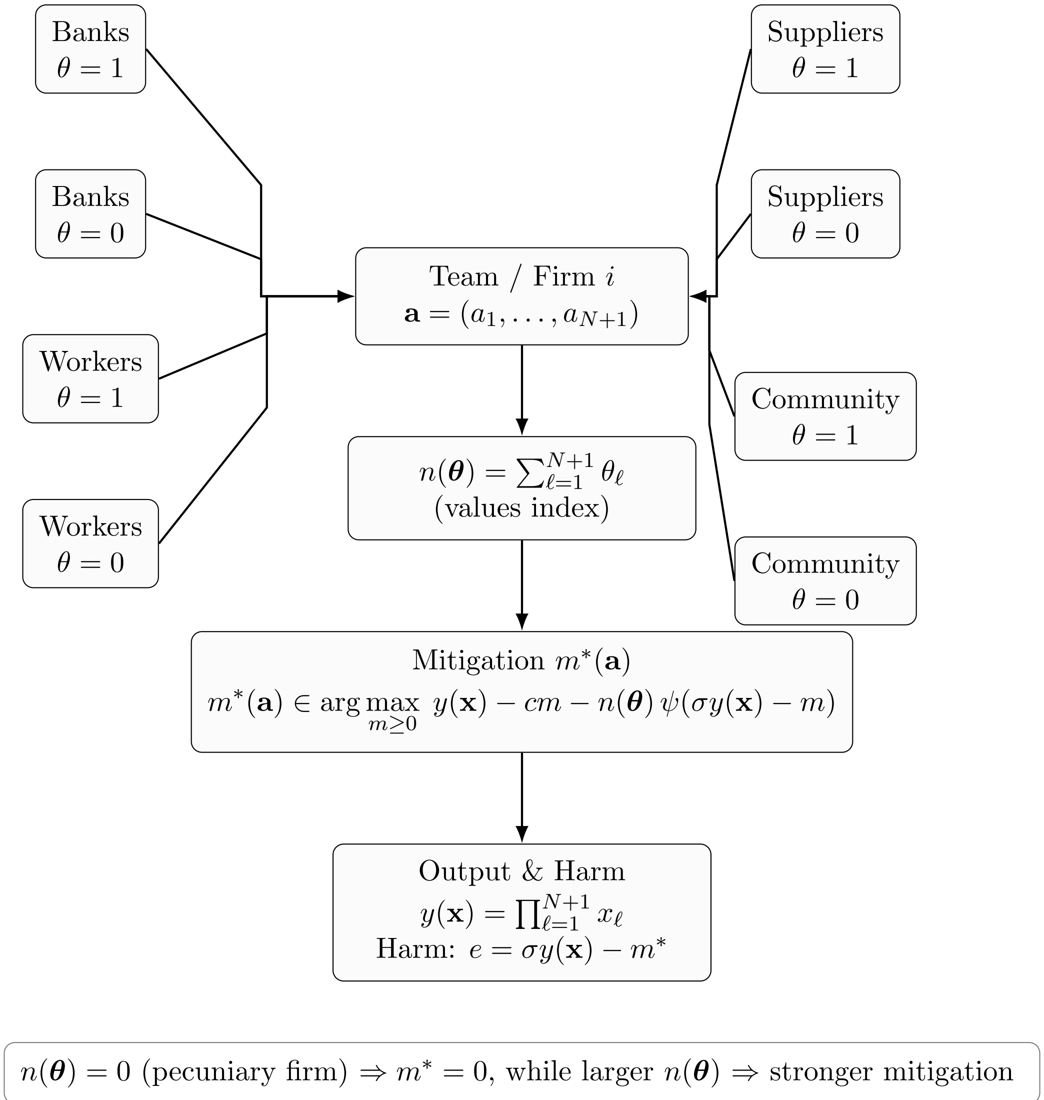

Figure 1: Extended schematic: banks and workers (left), suppliers and community (right) match with a firm, which determines the values index, mitigation, and resulting output and harm.

4.1 Joint surplus

From Equation 1, the surplus of any team can be written as

$$\Lambda(\mathbf{a}) = \Omega(y(\mathbf{x}), n(\boldsymbol{\theta})),$$

where

$$\Omega(y, n) = \max_{m \geq 0} \Big\{ y - cm - n\psi(\sigma y - m) \Big\}. \tag{3}$$

We impose assumptions to guarantee that mitigation is interior whenever $n \geq 1$, and that production is always valuable despite mitigation costs. Let $\underline{x}_\ell$ be the lowest skill of type ℓ and $\underline{\mathbf{x}} = (\underline{x}_1, \ldots, \underline{x}_{N+1})$ the least productive profile.

Assumption 1. *(i) (Interior mitigation) $\psi'(\sigma y(\underline{\mathbf{x}})) > c$. (ii) (Production valuable) $1 - c\sigma > 0$ and $\Omega(y(\underline{\mathbf{x}}), N) > 0$.*

4.2 Sorting on values

We first consider how stakeholders sort on values, holding productivity fixed. Adding a values-driven stakeholder is costly ($\Omega_n(y, n) < 0$), but mitigation is non-rival, so its cost is shared across all values-driven members. This implies clustering: values-driven agents prefer to join each other.

Let $(y_{-\ell}, n_{-\ell})$ denote the productivity and stakeholder-values index of a team without stakeholder (x_ℓ, θ_ℓ). Let $n^*_{-\ell}(x_\ell, \theta)$ be the equilibrium index of the team matched with stakeholder (x_ℓ, θ).

Lemma 1 (Concentration of values-driven stakeholders)**.** *Consider two type stakeholders of type ℓ with the same ability ($x_\ell = x_\ell$), then*

$$n^*_{-\ell}(x_\ell, 1) \;\geq\; n^*_{-\ell}(x_\ell, 0).$$

That is, conditional on the skill, values-driven stakeholders must join a team with a higher value-index than the equivalent pecuniary stakeholders. Formally, this result can be seen from the fact that $\Omega(y, n)$ is decreasing and strictly convex in n, which implies that a more extreme distribution of n results in higher aggregate surplus. Thus, it can never be the case that the values-driven stakeholder is matched with the firm with a lower stakeholder-values index than the equivalent pecuniary stakeholder. Otherwise, one can switch these two agents, which results in a more extreme distribution of n but does not affect the output in each team.

4.3 Productivity–values interdependence

Next we study how productivity and values interact. Output always raises surplus ($\Omega_y(y,n) > 0$), but the marginal value of productivity depends on the team's values index.

Intuitively, higher output creates more harm. Thus, productivity is worth less to teams with values-driven members than to purely pecuniary teams. Formally, by the envelope theorem,

$$\Omega_y(y,n) = 1 - n\psi'(\sigma y - m^*(y,n))\sigma. \tag{4}$$

For $n \geq 1$, the FOC implies $n\psi'(\sigma y - m^*(y,n)) = c$, so

$$\Omega_y(y,n) = 1 - c\sigma < \Omega_y(y,0) = 1. \tag{5}$$

Hence, the marginal value for output is only one for a purely pecuniary team ($n = 0$). On the other hand, any team with values-driven members discounts it to $1 - c\sigma$. This is because that with linear mitigation costs, the optimal mitigation is

$$m^*(y,n) = \sigma y - \xi_n^*, \qquad n \geq 1,$$

where ξ_n^* solves $n\psi'(\xi) = c$. Thus mitigation scales linearly with output, yielding the constant discount $1 - c\sigma$.

Robustness. With convex costs or disutility depending on harm rate, the same logic applies: $\Omega_y(y,n) \leq \Omega_y(y,0)$. In fact, $\Omega_{yn}(y,n) < 0$ so that higher n further lowers the marginal value of output. Our linear-cost specification keeps the model tractable while capturing the essence of the tradeoff.

Marginal value of skills Consider a stakeholder (x_ℓ, θ_ℓ) joining a team $(y_{-\ell}, n_{-\ell})$. The surplus is $\Omega(y_{-\ell}x_\ell,\, n_{-\ell}+\theta_\ell)$. The marginal contribution for a Values-driven stakeholder is thus

$$\frac{\partial}{\partial x_\ell}\Omega(y_{-\ell}x_\ell,\, n_{-\ell}+1) = (1-c\sigma)\, y_{-\ell}, \tag{6}$$

which rises with team productivity but is always discounted by $(1-c\sigma)$. On the other hand, for a pecuniary stakeholder, his marginal contribution is

$$\frac{\partial}{\partial x_\ell}\Omega(y_{-\ell}x_\ell,\, n_{-\ell}) = \begin{cases} (1-c\sigma)\, y_{-\ell}, & n_{-\ell} \geq 1, \\ y_{-\ell}, & n_{-\ell} = 0. \end{cases} \tag{7}$$

Thus, pecuniary skills are especially valuable in purely pecuniary teams. In other word, he can only avoid the discount if and only if he joins a pure pecuniary stakeholder team.

We thus now define an index that summarizes the joint effect of the team's productivity and values-driven preferences $(y_{-\ell}, n_{-\ell})$, which is given by

$$z_{-\ell}(y_{-\ell}, n_{-\ell}) \equiv \begin{cases} (1-c\sigma)\, y_{-\ell} & n_{-\ell} \geq 1 \\ y_{-\ell} & n_{-\ell} = 0 \end{cases}. \tag{8}$$

In other words, if a pecuniary stakeholder $(x_\ell, 0)$ is indifferent between a pecuniary team and a non-pecuniary team, then these two teams must have the same z-index, and the pecuniary team must be less productive. This is different for a values-driven stakeholder, whose marginal value is always ranked based on the productivity of the team $y_{-\ell}$ only. That is, regardless the stakeholder-values index of the team, his marginal value increases with the productivity of the team.

Lemma 2 (Positive assortative matching)**.** *(i) Among values-driven stakeholders ($\theta = 1$), higher skill x_ℓ implies matching with a higher-y team. (ii) Among pecuniary stakeholders ($\theta = 0$), higher skill x_ℓ implies matching with a team with higher z index*

Benchmark: productivity-only sorting This contrasts with the standard case where preferences are homogeneous and surplus is simply output. Then matching is positive assortative on ability alone: the i-th quantile of each type matches together, yielding firm output

$$y[i] = \prod_{\ell=1}^{N+1} x_\ell[i],$$

where $x_\ell[i]$ is the i-th quantile skill of type ℓ.

In our setting, by contrast, heterogeneity in non-pecuniary preferences and the interdependence between productivity and values mean that sorting must be characterized jointly on skills and preferences.

Summary. To summarize, optimal team formation features three key properties. First, values-driven stakeholders cluster together, since mitigation is non-rival and becomes cheaper in larger groups. Second, productivity and values are interdependent: additional output is worth less to teams that care about harm, so purely pecuniary teams place a higher marginal value on skills than values-driven ones. Third, this implies distinct sorting patterns. Values-driven stakeholders rank teams purely by productivity, while pecuniary stakeholders rank teams by a composite z-index that reflects both productivity and the absence of values-driven members. Relative to the benchmark of sorting only on skills, our framework highlights how heterogeneous preferences and the harm–output tradeoff fundamentally alter matching patterns in equilibrium.

5 Characterization

We now study how heterogeneous stakeholder values and productivity together affect the sorting. To fully characterize the matching outcome, we assume that the distribution of skills and values-driven preferences are independent and identically distributed throughout the rest of the paper for simplicity.

Assumption 2. *Skills and values are independently distributed. For any level of skill x_ℓ, the probability of being a values-driven stakeholder of type ℓ is λ_ℓ.*

Roadmap. We first illustrate equilibrium matching in a simple $N = 2$ setting. With balanced values-driven shares, teams fully segment into purely pecuniary or purely values-driven groups. When shares differ, however, segmentation breaks: high-skill agents remain separated, but lower-skill agents mix across preferences. We then extend the construction to general N via a sequential algorithm, and finally characterize transfers and premiums.

5.1 Illustrative example with $N = 2$

We now illustrate how equilibrium matches are constructed in the simplest case with $N = 2$ stakeholder types, say banks and workers. Because firms are pecuniary, Lemma 2 implies that more productive firms are always matched with higher z-index teams (Equation **??**). Thus we focus on how stakeholders form teams, with the firm assignment solved afterwards by positive assortative matching.

5.1.1 Balanced supply: full segmentation

Suppose first that the share of values-driven stakeholders is the same across both types, $\lambda_\ell = \lambda_{\ell'}$. In this case, the market segments cleanly. Values-driven stakeholders only match with values-driven ones, and pecuniary stakeholders only with pecuniary ones. Within each segment, sorting is positive assortative on skills, by Lemma 2.

Proposition 1 (Balanced supply)**.** *When $\lambda_\ell = \lambda_{\ell'}$, equilibrium matches are fully segmented: pecuniary agents form pure pecuniary teams and values-driven agents form pure values-driven teams, with positive assortative matching within each segment.*

This outcome is shown in Figure 2, where the dashed line denotes the z-index of pecuniary matches and the dotted line the z-index of values-driven matches. The values-driven line lies strictly below the pecuniary one due to the discount factor $(1 - c\sigma)$.

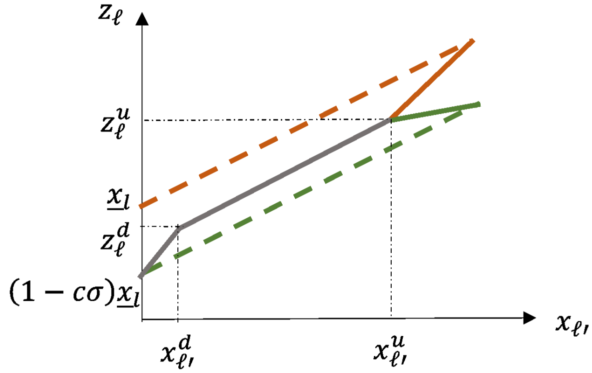

Figure 2: Balanced supply: pecuniary and values-driven stakeholders sort separately. The z-index of values-driven matches is lower due to the discount factor $(1 - c\sigma)$.

5.1.2 Unbalanced supply: segmentation at the top, mixing at the bottom

When $\lambda_\ell \neq \lambda_{\ell'}$, full segmentation cannot hold everywhere. Suppose $\lambda_\ell > \lambda_{\ell'}$. In this case, some values-driven stakeholders of type ℓ must match with pecuniary stakeholders of type ℓ' to clear the market.

Proposition 2 (Unbalanced supply). *When $\lambda_\ell > \lambda_{\ell'}$, there exists a cutoff $\hat{x}_{\ell'}$ such that: (i) for $x_{\ell'} \geq \hat{x}_{\ell'}$, pecuniary stakeholders of type ℓ' match only with pecuniary type-ℓ stakeholders (full separation at the top); (ii) for $x_{\ell'} < \hat{x}_{\ell'}$, pecuniary stakeholders of type ℓ' mix with values-driven type-ℓ stakeholders (mixing at the bottom).*

Relative to the balanced case, pecuniary stakeholders of type ℓ are scarce. This scarcity implies that keeping a purely pecuniary team must sacrifice productivity. At the top of the skill distribution, however, pecuniary stakeholders of type ℓ' still prefer pecuniary matches, since mitigation cost is relatively high. At the bottom, where productivity losses are smaller, pecuniary stakeholders of type ℓ' accept values-driven partners, leading to mixing.

Formally, define $\Psi_\ell(z)$ as the measure of type-ℓ stakeholders with effective z-index below z:

$$\Psi_\ell(z) \equiv \int_{\underline{x}}^{z/(1-c\sigma)} g_\ell^1(\tilde{x}_\ell)\, d\tilde{x}_\ell + \int_{\underline{x}}^{z} g_\ell^0(\tilde{x}_\ell)\, d\tilde{x}_\ell, \tag{9}$$

where g_ℓ^1 and g_ℓ^0 are the densities of values-driven and pecuniary stakeholders, respectively.

Positive assortative matching then implies

$$\Psi_\ell(\phi_\ell^m(x_{\ell'})) = F_{\ell'}(x_{\ell'}),$$

where $\phi_\ell^m(x_{\ell'})$ gives the z-index of the type-ℓ partner matched with stakeholder $x_{\ell'}$. The slope of ϕ_ℓ^m lies between the values-driven and pecuniary dashed lines in Figure 2, reflecting competition between types.

Interpretation Two key features emerge. First, high-skill stakeholders remain separated, as the cost of mixing is too high at the top. Second, at the bottom, mixing occurs: excess values-driven stakeholders of one type pair with pecuniary stakeholders of the other. Importantly, not all excess values-driven stakeholders end up at the very bottom; some mid-skill pecuniary stakeholders are also willing to mitigate when values-driven partners are sufficiently productive. This highlights how competition endogenously determines which part of the distribution bears the cost of mixing.

5.2 General case: sequential algorithm for $N > 2$

We now extend the construction to the case with N stakeholder types in the following sequential manner, where at each step, one type of stakeholder is added to an existing team, and the matching is determined by the same principles as in the $N = 2$ case.

Step 1: ordering of types. Label stakeholder types such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, where λ_ℓ is the share of values-driven stakeholders in type ℓ. Type 1 thus has the most values-driven stakeholders, while later types are increasingly "pecuniary."

Step 2: initial matching ($N = 2$). In the first stage, type 1 and type 2 match. This reduces to the $N = 2$ problem analyzed above: if $\lambda_1 = \lambda_2$, they segment cleanly into values-driven and pecuniary teams; if not, separation holds at the top but mixing occurs at the bottom. The relevant index for pecuniary stakeholders is their team z-index, while values-driven stakeholders sort by y.

Step 3: adding the next type fixing the team in eariler rounds. Once types 1 and 2 are matched, treat each pair as a "team." In the next stage, these teams are matched with type 3 stakeholders. Since type 3 has fewer values-driven members than the combined team of types 1–2, the same logic applies: at the top, separation holds, while at the bottom, some mixing occurs. Here, type 3 stakeholders are placed on the x-axis, and the relevant index for the team is its effective z-index.

Step 4: iteration. Repeat this procedure sequentially. At each stage τ, the team S_τ consisting of types $1, \ldots, \tau$ matches with stakeholders of type $\tau + 1$. By construction, type $\tau + 1$ is relatively more pecuniary, so the same separation–mixing logic applies. After matching, the team updates its characteristics:

$$y_{\tau+1} = y_\tau \cdot x^*_{\tau+1}(S_\tau), \qquad n_{\tau+1} = n_\tau + \theta^*_{\tau+1}(S_\tau).$$

Final step: matching with firms. After N rounds, each team contains all stakeholder types. Since firms are pecuniary by assumption, the final match is straightforward: firms sort positively with teams by productivity, i.e. x_{N+1} matches with the team z-index.

Proposition 3 (Sequential construction of equilibrium)**.** *Under Assumption 2, the equilibrium matching outcome can be constructed sequentially. At each stage τ, there exists a cutoff $x^u_{\tau+1}$ such that: (i) stakeholders above the cutoff match only within their own preference type (full separation at the top), (ii) stakeholders below the cutoff mix, with positive assortative matching between the team's z-index and stakeholder ability $x_{\tau+1}$. Conditional on mixing, values-driven stakeholders are matched first with values-driven teams.*

Evolution of values-driven teams. This sequential process implies two patterns. First, once a team becomes pecuniary ($n_\tau = 0$), it remains pecuniary in all subsequent stages, since later types are more pecuniary. Second, values-driven teams do not necessarily become more values-driven as they grow: in later stages, values-driven stakeholders are relatively scarce, so some matches add pecuniary members. In particular, the total measure of pecuniary teams is $(1-\lambda_1)$, the mass of pecuniary type-1 stakeholders.

5.3 Transfers and compensating differentials

We now turn to compensation. Stakeholders' utilities depend on both their skill and whether they are pecuniary or values-driven. Equation 2 implies that

$$\frac{\partial U_\ell(x_\ell,\theta)}{\partial x_\ell} = z^*_{-\ell}(x_\ell,\theta),$$

so the marginal gain from ability equals the effective productivity (z-index) of the matching team. Thus, unless the team is purely pecuniary, all stakeholders' marginal contributions are discounted by $(1-c\sigma)$.

Pecuniary stakeholders For pecuniary agents $(x_\ell,0)$, equilibrium utility coincides with the transfer they receive:

$$p_\ell(x_\ell,0) = U_\ell(x_\ell,0).$$

Hence, pecuniary compensation is uniquely pinned down by integration from the lowest skill:

$$U_\ell(x_\ell,0) = \int_{\underline{x}_\ell}^{x_\ell} z^*_{-\ell}(\tilde{x}_\ell,0)\,d\tilde{x}_\ell + U_\ell(\underline{x}_\ell,0).$$

All pecuniary stakeholders of type ℓ therefore receive the same fee, regardless of whether their team is matched with values-driven or pecuniary partners.

Values-driven stakeholders Values-driven stakeholders $(x_\ell,1)$ care both about transfers and about firm harm. Let $p_\ell(x_\ell,1|n)$ denote the fee when matched with a team of index n.

Their total utility is

$$U_\ell(x_\ell, 1) = p_\ell(x_\ell, 1|n) - \psi(\xi_n^*),$$

so the fee is

$$p_\ell(x_\ell, 1|n) = U_\ell(x_\ell, 1) + \psi(\xi_n^*).$$

If a stakeholder is indifferent between two teams with indices n and n', the fee difference exactly equals the difference in disutility from harm:

$$p_\ell(x_\ell, 1|n) - p_\ell(x_\ell, 1|n') = \psi(\xi_n^*) - \psi(\xi_{n'}^*).$$

This is a standard compensating differential: more harmful teams must pay higher wages to attract values-driven stakeholders.

Premiums and rents In the symmetric case (balanced supply), pecuniary stakeholders generally earn a premium relative to the compensating differential (Rosen 1986, Lavetti 2023), since they are especially valuable to firms. With unbalanced supply, however, some pecuniary and values-driven stakeholders of the same type may end up matched with teams of the same z-index. In this case, pecuniary stakeholders no longer capture additional rents: they are paid the same as otherwise identical values-driven stakeholders, apart from the compensating differential.

Proposition 4 (Premiums for pecuniary stakeholders)**.** *Pecuniary stakeholders of the most values-driven type (type 1) earn a positive ranking premium: their compensation reflects their higher value to firms. By contrast, pecuniary stakeholders in the mixing region ($x_\ell \leq x_\ell^u$) earn no additional premium beyond the compensating differential, since they are paid the same as values-driven counterparts matched with the same team.*

6 Impact of Values-Driven Preference Shocks

We now use the equilibrium characterization to study how an exogenous increase in the share of values-driven stakeholders affects matching patterns and firm harm. Formally, consider a shock that increases the measure of values-driven stakeholders of type ℓ (i.e., λ_ℓ). This lets us analyze whether stakeholders optimally remain with their existing firms (engagement) or reallocate to new firms (exit), and the resulting impact on harm.

Recall that a firm's harm depends on its productivity y and stakeholder-values index n:

$$e^*(y,n) = \begin{cases} \sigma y & n = 0 \quad \text{(pure pecuniary team)} \\ \xi_n^* & n \geq 1 \quad \text{(values-driven team)} \end{cases}.$$

When $n \geq 1$, harm depends only on n and not on y. Thus, the distribution of the stakeholder-values index across firms summarizes the first-order impact of preference shocks. Only for purely pecuniary firms ($n = 0$), harms depend on productivity, so this second channel is quantitatively minor. We therefore focus primarily on changes in the distribution of n.

6.1 Aggregate Impact

Because values-driven stakeholders cluster, the distribution of firms by stakeholder-values index is pinned down mechanically by the vector $(\lambda_1, \ldots, \lambda_N)$. For example, with $N = 2$, the measure of firms with two values-driven stakeholders is λ_2 (the scarcer type), and the measure with exactly one values-driven stakeholder is $\lambda_1 - \lambda_2$. Increasing λ_2 by δ raises the measure of fully values-driven firms by δ and reduces the measure of partially values-driven firms by δ, leaving pecuniary firms unchanged. This logic generalizes to any N.

Proposition 5 (Aggregate effect)**.** *The share of firms with stakeholder-values index ℓ equals $\lambda_\ell - \lambda_{\ell+1}$ for $\ell = 1, \ldots, N$, and the share of purely pecuniary firms equals $1 - \lambda_1$. If the ordering of λ_ℓ is unchanged, then increasing λ_ℓ by δ raises the share of firms with index ℓ by δ and lowers the share with index $\ell - 1$ by δ.*

6.2 Micro-Level Impact

We now turn to the impact at the individual stakeholder level. Suppose a small mass δ of type-ℓ stakeholders switch from pecuniary to values-driven ("treated stakeholders"). Their effect depends on whether they stay in their original firm (engagement) or reallocate to another firm (exit). We show that engagement dominates at the bottom of the distribution, exit occurs at the top, and in the middle treated stakeholders trigger additional harm reduction through a trickle-down effect.

6.2.1 Exit vs. Engagement

Proposition 6 (Exit vs. Engagement)**.** *Suppose $\ell \geq 2$ and the share of values-driven type-ℓ stakeholders increases by a small $\delta > 0$ with $\lambda_\ell + \delta < \lambda_{\ell-1}$. Then there exists a cutoff firm size x^d_{N+1} such that: (i) treated stakeholders below the cutoff remain with their original firm (engagement); (ii) treated stakeholders above the cutoff exit to smaller firms with a higher stakeholder-values index.*

Intuition. The cutoff arises because the separation threshold in Figure 2 shifts downward after the shock. At the top, firms prefer pecuniary partners, so newly values-driven stakeholders exit to smaller values-driven firms. At the bottom, firms are already mixing, so treated stakeholders can stay and reduce harm through engagement.

Engagement at the Bottom At the bottom of the distribution, pecuniary and values-driven stakeholders of type ℓ are ranked equally once matched with at least one values-driven partner. Hence, when a low-ability stakeholder turns values-driven, his equilibrium match does not change. He remains with the same firm, which now has one additional values-driven member, lowering harm from $\xi^*_{\ell-1}$ to ξ^*_ℓ.

Proposition 7 (Engagement impact)**.** *All treated stakeholders below the cutoff stay in their original firm. Each reduces harm by $(\xi^*_\ell - \xi^*_{\ell-1})$, so the total impact in this region is $(\xi^*_\ell - \xi^*_{\ell-1})\delta$.*

Exit and Displacement at the Top At the top, treated stakeholders exit to more values-driven firms. This has two consequences: (i) they may have no direct effect if the firm already had a values-driven stakeholder, and (ii) they displace untreated values-driven stakeholders downward in the distribution. Thus, while treated top stakeholders themselves may not reduce harm, they set off reallocations that do.

Trickle-Down Effect in the Middle The displacement mechanism implies that additional harm reduction must occur in the middle of the distribution. Treated high-ability values-driven stakeholders crowd out lower-ability untreated ones, who then join smaller firms and make those firms more values-driven. This generates more harm reduction than would be achieved by treated stakeholders alone.

Proposition 8 (Trickle-down effect)**.** *(i) At the bottom ($x_\ell \leq x_\ell^d$), the harm reduction exactly equals the measure of treated stakeholders. (ii) In a middle region ($x_\ell^d < x_\ell < x_\ell^u$), the harm reduction exceeds the measure of treated stakeholders, as displaced untreated stakeholders also reduce harm. (iii) At the top ($x_\ell \geq x_\ell^u$), some treated stakeholders have no direct effect, as they join firms whose harm does not change.*

Interpretation. Aggregate harm reduction equals exactly $\delta(\xi_\ell^* - \xi_{\ell-1}^*)$ (Proposition 5). But the distribution of this impact across firms is uneven: the bottom experiences direct engagement effects, the top may see little change, and the middle gains disproportionately through trickle-down.

Summary. The effect of values-driven preference shocks on the aggregate harm can be summarized by the change in λ_ℓ. At the micro level, treated stakeholders at the bottom engage and reduce their firm's harm directly, while treated stakeholders at the top often exit without direct effect but trigger trickle-down reallocations. As a result, the aggregate impact is realized disproportionately in the middle of the distribution, even though the shocks are uniformly distributed.

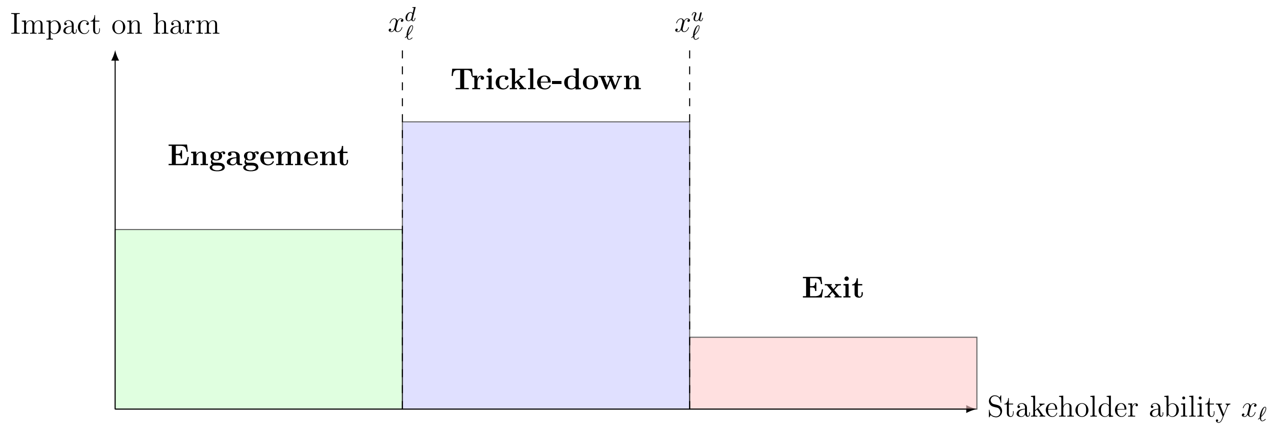

Figure 3: Schematic of micro-level impact of a values-driven preference shock. At the bottom ($x_\ell \leq x_\ell^d$), treated stakeholders stay and reduce harm directly (engagement). In the middle ($x_\ell^d < x_\ell < x_\ell^u$), displacement of untreated stakeholders generates additional harm reduction (trickle-down). At the top ($x_\ell \geq x_\ell^u$), treated stakeholders reallocate but often have little direct impact (exit).

7 Calibration

We now calibrate the model to quantify the magnitude of the trickle-down effect in the data. The exercise focuses on the 500 most carbon-intensive publicly listed firms, primarily in the power sector, and two types of stakeholders: banks and workers. The calibration is informed by recent empirical evidence on the effect of values-driven financiers on corporate emissions (Kacperczyk and Peydró 2022; Duchin, Gao, and Xu 2022; Akey and Appel 2019; Hartzmark and Shue 2022). These studies show that when values-driven banks reallocate their lending, the direct effect on the new firms is limited, while firms that lose financing often increase emissions. This pattern is consistent with our model: the local effect of exit is small, and the aggregate effect operates primarily through spillovers.

Relative shares of values-driven stakeholders. Survey evidence suggests that values-driven preferences are more prevalent among workers than among banks. According to an IBM survey of 14,000 households, 33% of workers accepted values-driven jobs at an average wage discount of 28% (see also Krueger, Metzger, and Wu 2021). By contrast, Kacperczyk and Peydró (2022) report that only 7% of bank loans go to values-driven firms. Consistent with our model, differences in λ across stakeholder groups are reflected in compensation: banks show only small interest-rate differentials, while workers face large wage differentials in values-driven firms.

Parameters. We set the worker share of values-driven preferences to $\lambda_1 = 33\%$ (IBM survey). Worker talent distribution follows Branikas et al. (2022) with $\gamma_1 = -0.4$ and support $[0.08, 0.3]$. The share of values-driven banks is $\lambda_2 = 7\%$ at $t = 0$, with support $[1.1, 1.4]$, based on Kacperczyk and Peydró (2022). Bank talent distribution parameters $(\gamma_2, \underline{x}_2, \overline{x}_2)$ are chosen to match asset and debt distributions from bank loan data, with bank assets given by $x_2 x_3$ and debt by $(x_2 x_3 - x_3)$. Firm productivity distribution, from Branikas et al. (2022), has $\gamma_3 = 5$ and support $[5000, 100000]$.

For emissions, we set $\sigma = 5000$ using the ratio of carbon emissions to firm revenues from Trucost. Mitigation cost is calibrated to carbon capture surveys with $c = 0.00008$, implying $c\sigma = 0.04$. Remaining parameters $\kappa, \rho, \xi_1^*, \xi_2^*$ are chosen to fit the production–emissions relationship (Figure 4).[3]

Results. We consider an increase in the share of values-driven banks from $\lambda_2 = 7\%$ at $t = 0$ to $\lambda_2 = 15\%$ at $t = 1$, i.e. an 8% treatment shock.

Effect on firms. Figure 5 shows that at the bottom of the distribution, exactly 8% of firms transition from grey to dark green, equal to the treated share. In the middle (firms ranked between 400–600), more than 8% of firms transition, because many teams that were mixed (one values-driven bank) before the shock become fully values-driven afterwards. Since the aggregate impact must equal 8%, this implies that the impact at the top is less than 8%.

Impact from the stakeholder's viewpoint. We distinguish between BG banks (brown at $t = 0$, green at $t = 1$) and GG banks (green in both periods). Figure 6 shows the share of banks with measurable impact, defined as making their matched firm greener at $t = 1$. If all treated banks engaged their original firms, 8/15 of green banks (53%) would have impact.

[3]Specifically, $e = \xi_2^* = 5 \times 10^6$ when $n = 2$; $e = \xi_1^* = 10^7$ when $n = 1$; and $e = \sigma Y$ when $n = 0$, giving $\sigma = 5000$. The system $n\psi'(\xi_n^*) = c$ with $\psi(\xi) = \frac{\kappa}{1+\rho}\xi^{1+\rho}$ yields $\kappa = 8 \times 10^{-12}$, $\rho = 1$.

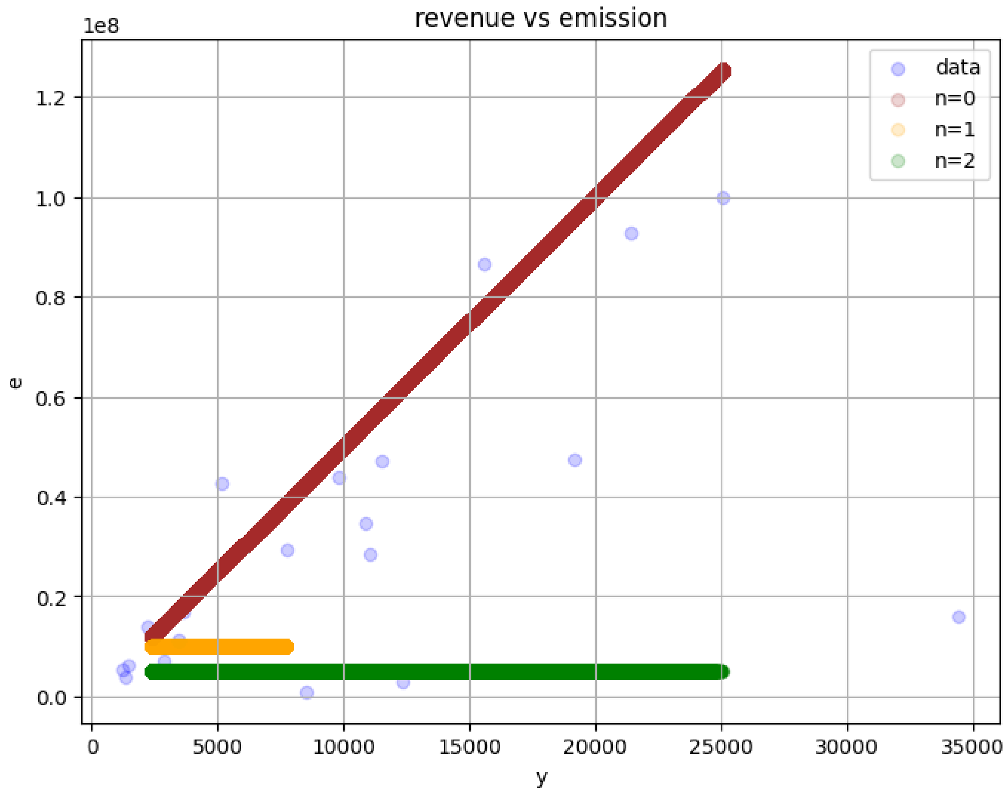

Figure 4: Calibrated production–emission relation.

However, only the bottom banks engage, so only their realized impact is exactly 53%. At the top, some treated BG banks have no direct effect: instead, they displace GG banks, who then make more mid-range firms become greener. This reallocation generates the trickle-down impact.

Interpretation. The calibration illustrates our theoretical mechanisms in the data. At the bottom of the distribution, treated banks remain with their original firms, lowering harm directly (engagement). At the top, many treated banks exit without immediate impact, but in doing so displace already green banks, who then shift mid-tier firms into greener matches. This displacement generates the trickle-down impact, whereby the aggregate reduction in emissions exceeds the direct contribution of treated banks alone. Thus, the calibration confirms the model's central prediction: while local effects of exit appear limited, the spillover reallocations are quantitatively significant and drive most of the aggregate reduction in harm.

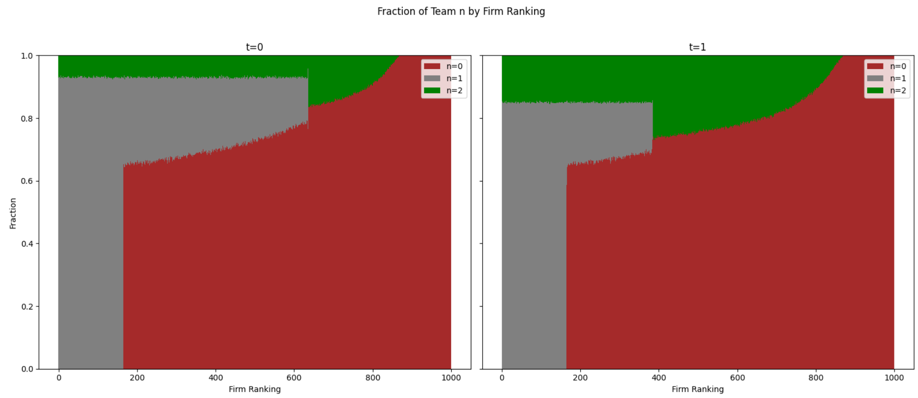

Figure 5: Firm matching outcomes before vs. after the shock.

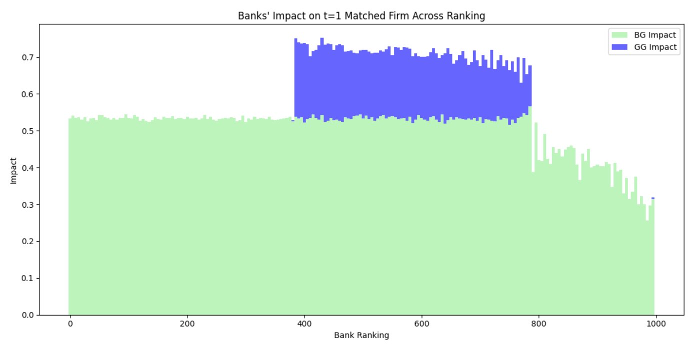

Figure 6: Share of banks with impact at $t = 1$. BG = treated, GG = already green.

8 Conclusion

This paper revisits the long-standing debate over exit versus engagement by modeling values-driven stakeholders in a general equilibrium framework. Our analysis highlights that the aggregate impact of exit has been systematically understated. While exit by highly productive stakeholders has limited direct effect—since they reallocate to firms that already mitigate—its influence trickles down through equilibrium reallocation. Less-productive stakeholders, displaced by these movements, shift to new firms, thereby inducing additional mitigation responses. Engagement remains an important channel, but our results demonstrate that exit, often dismissed as blunt or ineffective, can generate substantial spillovers that amplify its impact.

A calibration to data on banks and workers underscores the empirical relevance of these dynamics, suggesting that exit may play a larger role in shaping corporate environmental and social outcomes than previously thought. Beyond clarifying the conditions under which exit and engagement differ, our framework provides a tool for understanding the interplay between stakeholder values and firm behavior in competitive markets.

Future work could extend this analysis to heterogeneous values intensities across stakeholders, dynamic responses over time, and institutional settings where the balance between exit and engagement is mediated by governance structures. More broadly, our findings suggest that policy debates and corporate governance reforms should not underestimate the power of exit, especially when considered through the lens of spillovers and general equilibrium effects.

References

Acemoglu, Daron, Ufuk Akcigit, and William R Kerr (2016). "Innovation network". In: *Proceedings of the National Academy of Sciences* 113.41, pp. 11483–11488.

Akerlof, George A and Rachel E Kranton (2000). "Economics and identity". In: *The quarterly journal of economics* 115.3, pp. 715–753.

Akey, Pat and Ian Appel (2019). "Environmental externalities of activism". In: *Available at SSRN 3508808*.

Bénabou, Roland and Jean Tirole (2006). "Incentives and prosocial behavior". In: *American economic review* 96.5, pp. 1652–1678.

Besley, Timothy and Maitreesh Ghatak (2005). "Competition and incentives with motivated agents". In: *American economic review* 95.3, pp. 616–636.

Branikas, Ioannis et al. (2022). *Sustainability Preferences of Talented Employees*. Tech. rep. SSRN Working Paper.

Broccardo, Eleonora, Oliver Hart, Luigi Zingales, et al. (2022). "Exit versus voice". In: *Journal of Political Economy* 130.12, pp. 3101–3145.

Chiappori, Pierre-André, Robert McCann, and Brendan Pass (2016). "Multidimensional matching". In: *arXiv preprint arXiv:1604.05771*.

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2018). "Bidimensional matching with heterogeneous preferences: education and smoking in the marriage market". In: *Journal of the European Economic Association* 16.1, pp. 161–198.

Dimson, Elroy, Oğuzhan Karakaş, and Xi Li (2015). "Active ownership". In: *The Review of Financial Studies* 28.12, pp. 3225–3268.

Duchin, Ran, Janet Gao, and Qiping Xu (2022). "Sustainability or greenwashing: Evidence from the asset market for industrial pollution". In: *Available at SSRN 4095885*.

Dupuy, Arnaud and Alfred Galichon (2014). "Personality traits and the marriage market". In: *Journal of Political Economy* 122.6, pp. 1271–1319.

Ellingsen, Tore and Magnus Johannesson (2008). "Pride and prejudice: The human side of incentive theory". In: *American economic review* 98.3, pp. 990–1008.

Ellison, Glenn and Edward L Glaeser (1997). "Geographic concentration in US manufacturing industries: a dartboard approach". In: *Journal of political economy* 105.5, pp. 889–927.

Epple, Dennis, Thomas Romer, and Holger Sieg (2001). "Interjurisdictional sorting and majority rule: an empirical analysis". In: *Econometrica* 69.6, pp. 1437–1465.

Gabaix, Xavier and Augustin Landier (2008). "Why has CEO pay increased so much?" In: *The Quarterly Journal of Economics* 123.1, pp. 49–100.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings". In: *Journal of political economy* 118.3, pp. 536–598.

Hartzmark, Samuel M and Kelly Shue (2022). "Counterproductive sustainable investing: The impact elasticity of brown and green firms". In: *Available at SSRN 4359282*.

Hatfield, John William and Scott Duke Kominers (2015). "Multilateral matching". In: *Journal of Economic Theory* 156. Computer Science and Economic Theory, pp. 175–206. ISSN: 0022-0531.

Heinkel, Robert, Alan Kraus, and Josef Zechner (2001). "The effect of green investment on corporate behavior". In: *Journal of financial and quantitative analysis* 36.4, pp. 431–449.

Hirschman, Albert O (1972). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Harvard university press.

Hong, Harrison, Neng Wang, and Jinqiang Yang (2021). *Welfare consequences of sustainable finance*. Tech. rep. National Bureau of Economic Research.

Huber, Kilian (2023). "Estimating general equilibrium spillovers of large-scale shocks". In: *The Review of Financial Studies* 36.4, pp. 1548–1584.

Kacperczyk, Marcin T and José-Luis Peydró (2022). "Carbon emissions and the bank-lending channel". In: *Available at SSRN 3915486*.

Krueger, Philipp, Daniel Metzger, and Jiaxin Wu (2021). "The sustainability wage gap". In: *Swedish House of Finance Research Paper* 20-14, pp. 21–17.

Lavetti, Kurt (2023). "Compensating wage differentials in labor markets: Empirical challenges and applications". In: *Journal of Economic Perspectives* 37.3, pp. 189–212.

Lindenlaub, Ilse (2017). "Sorting multidimensional types: Theory and application". In: *The Review of Economic Studies* 84.2, pp. 718–789.

Oehmke, Martin and Marcus M Opp (2023). "A theory of socially responsible investment". In: *Swedish House of Finance Research Paper* 20-2.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor (2021). "Sustainable investing in equilibrium". In: *Journal of Financial Economics* 142.2, pp. 550–571.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski (2021). "Responsible investing: The ESG-efficient frontier". In: *Journal of Financial Economics* 142.2, pp. 572–597.

Rosen, Sherwin (1986). "The theory of equalizing differences". In: *Handbook of labor economics* 1, pp. 641–692.

Sattinger, Michael (1979). "Differential rents and the distribution of earnings". In: *Oxford Economic Papers* 31.1, pp. 60–71.

Tervio, Marko (2008). "The difference that CEOs make: An assignment model approach". In: *American Economic Review* 98.3, pp. 642–68.

Tiebout, Charles M (1956). "A pure theory of local expenditures". In: *Journal of political economy* 64.5, pp. 416–424.

A Appendix

A.1 Omitted Proofs

A.1.1 Proof for Lemma 1

Proof. Obverse that $\Omega(y,n) = \max_{e\leq\sigma y}\{y - n\psi(e) - c(\sigma y - e)\}$ is decreasing and convex in n. This is because that $f(n,e|y) \equiv y - n\psi(e) - c(\sigma y - e)$ is linear in n and thus $\Omega(y,n) = \max_e f(n,e|y)$ is (strictly) convex in n. This holds for a general cleaning cost function. With the specified linear cost, it can further be reduced to

$$\Omega(y,n) = (1 - c\sigma)y + \chi(n,y),$$

where $\chi(y,n) \equiv \max_{e\geq c\sigma y}\{ce - n\psi(e)\} = c\xi_n^* - n\psi(\xi_n^*)$ for any $n \geq 1$, and $\chi(y,0) = c\sigma y$. This further implies that $\chi_n < 0$ and $\chi(y,n)$ is convex in n ($\chi_{nn} > 0$). The property of $\chi(y,n)$ can be summarized by the Lemma below. [4]

Lemma 3. $\chi(y,n) - \chi(y,n+1)$ *decreases in* n $\forall n$*, and* $\chi(y,n) - \chi(y,n+1)$ *is independent of* y *for* $n \geq 1$*, and increasing in* y *only when* $n = 0$.

We now prove this result by contradiction. Suppose the the values-driven index of the team for the values-driven agent's $(x_i, 1)$ is lower that the one of an otherwise identical pecuniary agent x_j, where $n_{-i} < n_{-j}$. We now show that the profitable deviation exists by switching their team. Intuitively, as both agents have the same ability, switching their teams do not affect the team productivity; however, since $\chi(y,n)$ is convex in n, switching results in more extreme value of n and thus increase total surplus.

[4] Our assumption 1 implies that it's optimal for any team to mitigate as long as there is one values-driven stakeholder. More generally, similar properties hold as long as the interior solution exists for any $n \geq \hat{n}$. In this case, $\chi(y,n) = c\sigma y$ $\forall n < \hat{n}$.

Formally, given $x_i = x_j = x$, the total surplus after switching yields

$$\{\Omega(y_{-i}x_j, n_{-i}) + \Omega(y_{-j}x_i, n_{-j}+1)\} - \{\Omega(y_{-i}x_i, n_{-i}+1) + \Omega(y_{-j}x_j, n_{-j})\}$$
$$= \{\chi(y_{-i}x_j, n_{-i}) + \chi(y_{-j}x_i, n_{-j}+1)\} - \{\chi(y_{-i}x_i, n_{-i}+1) + \chi(y_{-j}x_j, n_{-j})\} > 0$$

where the inequality uses the fact when $n_j > n_i > 0$, Lemma 3 implies that

$$\chi(y_{-i}x, n_{-i}) - \chi(y_{-i}x, n_{-i}+1) > \chi(y_{-j}x, n_{-j}) - \chi(y_{-j}x, n_{-j}+1).$$

What is left to show is when $n_i = 0$,and $n_{-j} > 0$, in this case, we have

$$\begin{aligned}\chi(y_{-i}x, 0) - \chi(y_{-i}x, 1) &> \chi(y_{-i}x, n_{-j}) - \chi(y_{-i}x, n_{-j}+1)\\ &= \chi(y_{-j}x, n_{-j}) - \chi(y_{-i}x, n_j+1)\end{aligned}$$

□

A.1.2 Proof for Lemma 2

Proof. A stakeholder's problem can be rewritten as choosing his team optimally, by taking as given the composition of the team which consists of all types of stakeholders (excluding his own type) and the total equilibrium utilities of agents in the team, which is denoted by $\mathit{\Pi}(y_{-\ell}, n_{-\ell}) \equiv \Sigma_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'})$. Hence, his optimization problem yields

$$U_\ell(x_\ell, \theta_\ell) = \max_{(y_{-\ell}, n_{-\ell})} \Omega(y_{-\ell}x_\ell, n_{-\ell} + \theta_\ell) - \mathit{\Pi}(y_{-\ell}, n_{-\ell})\Sigma_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'}).$$

Since Equation **??** implies complementarity between values-driven agent $(x_\ell, 1)$ and $y_{-\ell}$, hence, by the monotonic comparative statics, a values-driven agent with higher ability must choose a team with a higher productivity than a values-driven agent with lower ability. Similarly, Equation **??** implies complementarity between pecuniary agent $(x_\ell, 0)$ and $z_{-\ell}$; hence, a more skilled pecuniary agent must choose a team with a higher z-index. □

A.1.3 Proof for Proposition 2

Proof. Clearly, conditional on preference, PAM is the stable outcome due to the standard complementarity. Hence, what is left to show is that it is not optimal for agents to match across markets. Consider a values-driven stakeholder $(x_i, 1)$ a pecuniary stakeholder $(x_j, 0)$ switch teams, where $(y^*_{-i}, N-1)$ and $(y^*_{-j}, 0)$ represent their original team. The total surplus after switches yields

$$
\begin{aligned}
&\Omega(y^*_{-i}x_j, N-1) + \Omega(y^*_{-j}x_i, 1) \\
=&(1-c\sigma)\left(y^*_{-i}x_j + y^*_{-j}x_i\right) + \chi(y^*_{-i}x_i, N-1) + \chi(y^*_{-j}x_j, 1) \\
<&(1-c\sigma)\left(y^*_{-i}x_j + y^*_{-j}x_i\right) + \chi(y^*_{-i}x_i, N) + \chi(y^*_{-j}x_j, 0) \\
=&\Omega(y^*_{-i}x_i, N) + \Omega(y^*_{-j}x_j, 0),
\end{aligned}
$$

where the first equality uses the fact that $\chi(y, n)$ is independent of n for $n \geq 1$.

The second inequality uses the fact that, under PAM and full separation $y^*_{-i} \geq y^*_{-j}$ iff $x_i \geq x_j$. Hence, $\left(y^*_{-i}x_j + y^*_{-j}x_i\right) \leq y^*_{-i}x_i + y^*_{-j}x_j$; In other words, there is no productivity distortion in $(y_{-\ell}, x_\ell)$. That is, intuitively, there is no gain in aggregate surplus when matching across markets. There is, however, a cost of doing so, as $\chi(y, n)$ is convex in n, as we have for any $\hat{n} > 0$,according to Lemma 3,

$$\chi(y_j, 0) - \chi(y_j, 1) > \chi(y_j, N-1) - \chi(y_j, N) = \chi(y_i, N-1) - \chi(y_i, N).$$

□

A.1.4 Proof for Proposition 3

Proof. Observe that the sequential ordering implies the following properties: (1) for any pecuniary team at period τ, their team remains pecuniary after matches. That is, if $n_\tau = 0$, then $n_{\tau+1} = 0$. Intuitively, this is because that the stakeholders at the later periods, by construction, are more pecuniary. (2) given any team $S_\tau = (y_\tau, n_\tau)$,where $n_\tau \geq 1$, we have $n_{\tau+1} = n_\tau + 1$ if

$z_\tau(y_\tau, n_\tau) \geq z^u_{\tau+1}$, and conditional on $y_{\tau,}$ $n_{\tau+1} = n_\tau + 1$ iff $n_\tau \geq \hat{n}_\tau$. This is because that, in the mixing regions, conditional on z_τ, only the relative values-driven team get another values-driven stakeholder.

As a result, no matter for values-driven or pecuniary team, the evolution of their z index can be expressed as $z_{\tau+1} = z_\tau x^*_{\tau+1}(S_\tau)$,[5] and $n_{\tau+1} = n_\tau + \theta^*_{\tau+1}(S_\tau)$. Let $X^*_\tau(S_\tau)$ and $N^*_\tau(S_\tau)$ represent the optimal productivity and values-driven index chosen by the team S_τ from period τ to period N.

$$\Omega\left\{(y_\tau, n_\tau), (x_{\tau+1}, \theta_{\tau+1})\right\} = z_\tau(y_\tau, n_\tau)\, x_{\tau+1} X^*_{\tau+1}(S_{\tau+1}) + \chi\left(\left(n_\tau + \theta_{\tau+1} + N^*_{\tau+1}(S_{\tau+1})\right)\right).$$

Given that $X^*_{\tau+1}(S_{\tau+1})$ is monotonic in $z_{\tau+1}$ and $N^*_{\tau+1}(S_{\tau+1})$ is monotonic in $n_{\tau+1}$, it is sufficient to show that the matching outcome maximizes the product of $z_{\tau+1} = z_\tau x_{\tau+1}$ and the dispersion of $n_{\tau+1}$ at period τ. In other words, we now show that the matching is stable given any period τ. Since our construction satisfies Lemma **??**, conditional on the preference, the matching is stable.

What is left to show is there is no profitable deviation for stakeholders to match across types. Consider first the case where a values-driven stakeholder i considers to switch with a pecuniary stakeholder j,whose team has values-driven index $n_{-j} > 0$. That is, it must be the case where $x_j < x^d_\tau$

$$\begin{aligned}
&\tilde{\Omega}\left((y_{-i}, n_{-i}), (x_i, 1)\right) + \tilde{\Omega}\left((y_{-j}, n_{-j}), (x_j, 0)\right)\\
&-\left\{\tilde{\Omega}\left((y_{-i}, n_{-i}), (x_j, 0)\right) + \tilde{\Omega}\left((y_{-j}, n_{-j}), (x_i, 1)\right)\right\}\\
=&(1 - c\sigma)\left\{y_{-i}x_i + y_{-j}x_j - (y_{-i}x_j + y_{-j}x_i)\right\} + \left\{\chi(n_{-i} + 1) + \chi(n_{-j}) - (\chi(n_{-i}) + \chi(n_{-j} + 1))\right\} \geq 0,
\end{aligned}$$

The first term is positive, as by construction, $y_{-i} \geq y_{-j}$ iff $x_i \geq x_j$ given that $(x_j, 0)$ is in the mixing region. The second term is also positive as $\chi(n)$ is convex. Next, consider the case where $n_{-j} = 0$. In this case, the loss is even higher as both teams have to mitigate. □

[5] Importantly, this is not true if property (1) does not hold. This is because that, if a pecuniary team receives a values-driven stakeholder a period τ', then $z_{\tau+1} = z_\tau(1 - c\sigma)x^*_{\tau+1}(S_\tau)$.

A.1.5 Proof for Proposition 4

Proof. Since our equilibrium implies that if $x_\ell < x_\ell^d$, then the pecuniary stakeholder will be mixing between pecuniary and values-driven stakeholders with the team with same $z_{\tau-1}$ at period ℓ. As a result, they will have the same z_τ after the matches and since $\hat{x}_\tau$ increases in τ, the productivity of their sequential matching outcome remains the same. Hence, $z^*_{-\ell}(x_\ell, 0) = z^*_{-\ell}(x_\ell, 1)$. $\square$