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

Briana Chang† and Harrison Hong‡

October 2025

Abstract

We develop a general equilibrium model to study whether exit or engagement by values-driven stakeholders is more effective in mitigating firm harm. Firms and multi-type stakeholders form relationships that trade off production complementarities and mitigation costs. Using a novel characterization of a multi-sided, multi-dimensional matching equilibrium, we show that a values shock induces high-productivity stakeholders to exit toward firms that already mitigate, while low-productivity stakeholders optimally engage. Although exit has limited direct effects, it generates equilibrium reallocation for untreated stakeholders, inducing additional firms to mitigate. A calibration shows that, once these spillovers are accounted for, the aggregate impact of exit has been understated relative to engagement.

*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 consumer boycotts of controversial brands illustrate how values-driven stakeholders increasingly pressure firms to mitigate environmental and social harm. Should such stakeholders attempt to change firms by staying and engaging, or by exiting and reallocating? This question originates with Hirschman 1972 and remains central to debates in economics and policy. A growing literature argues that engagement dominates exit (such as Broccardo, Hart, Zingales, et al. 2022 and Dimson, Karakaş, and Li 2015), and exiting stakeholders often reallocate toward firms that are already socially responsible, which seems to fail to induce meaningful change. Yet exit continues to be widely used in practice by investors, workers, and consumers,[1] raising the puzzle of why stakeholders persist in strategies that appear ineffective.

This paper argues that the perceived weakness of exit reflects a narrow, firm-level perspective. When evaluated in general equilibrium, exit can generate effects that engagement cannot replicate. The central distinction is that engagement operates locally—within an existing relationship—whereas exit could trigger reallocation through market competition. Through this channel, exit can influence firms that are not directly affected by treated stakeholders.[2]

We develop a tractable framework in which firms form productive relationships with multiple types of stakeholders—such as workers, banks, and suppliers—who differ in both productivity and values. Values-driven stakeholders experience disutility from harm generated by the firms with which they are matched,[3] creating incentives for firms to mitigate environmental or social harm when such stakeholders are present. Stakeholder–firm relationships, output, the

[1] Prominent examples include Norges Bank's 2015 divestment from coal companies and a contemporaneous wave of large banks committing to divest from carbon-intensive firms. More broadly, divestment movements have received substantial support: the Global Fossil Fuel Commitments Database reports that 1,731 institutions have committed to divest approximately 40.76 trillion in assets as of December 22, 2025.

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

[3] Bonnefon et al. 2025 provides evidence of this type of preference, which is also refered as value-alignment in the portfolio choice literature.

mitigation of environmental or social harm, as well as profit-sharing, are jointly determined competitively. This structure allows us to study how a values shock leads stakeholders to either optimally sever relationships (exit) or remain and share mitigation costs (engage), taking into account not only direct effects but also the equilibrium reallocation spillovers that arise through market competition.

The model delivers a sharp characterization of equilibrium matching. Two forces play a central role. First, mitigation is non-rival within the firm: once undertaken, its cost is shared among all values-driven stakeholders in the team. This creates incentives for values-driven stakeholders to cluster, rather than spread evenly across firms. Second, productivity interacts with values because higher output generates more harm. As a result, productivity is less valuable in firms with values-driven stakeholders, where additional output must be partially offset by increased mitigation. Together, these forces generate systematic heterogeneity across firms in both mitigation behavior and stakeholder composition.

In equilibrium, this interaction leads to segmentation at the top of the productivity distribution. Above a critical productivity threshold, firms sort into two distinct groups: firms that hire exclusively values-driven stakeholders and undertake extensive mitigation, and firms that hire exclusively pecuniary stakeholders and do not mitigate. Below this threshold, full separation is not feasible. Some firms strategically adopt mitigation in order to attract more productive values-driven stakeholders, leading to mixed teams.

These equilibrium properties are crucial for understanding the choice between exit and engagement following a values shock. For highly productive stakeholders, inducing mitigation at top firms is too costly: such firms optimally avoid values-driven stakeholders, and any attempt at engagement would require large compensation transfers. Consequently, highly productive stakeholders optimally exit, reallocating toward smaller firms that host many values-driven stakeholders. At the other end of the productivity distribution, stakeholders face a different trade-off. Because lower-productivity firms are indifferent between stakeholders with identical productivity but different values, a values shock for these stakeholders does not destabilize existing relationships. Instead, such stakeholders optimally engage, remaining in their firms

and sharing the costs required to increase mitigation.

At the firm level, exit by highly productive stakeholders often appears ineffective: destination firms may already mitigate, and the firms they leave may continue operating with pecuniary stakeholders. This observation underlies much of the skepticism toward exit in the existing literature. However, this firm-level view misses a critical channel. When highly productive values-driven stakeholders exit, they displace less productive values-driven stakeholders at their destination firms. These displaced stakeholders then reallocate to other firms, where their presence induces mitigation that would not otherwise occur. Through this process, exit generates reallocation spillovers that propagate through the economy, affecting firms far removed from the original exit decision.

As a result, exit and engagement differ fundamentally in their aggregate effects. Engagement delivers direct, localized change within a firm but remains confined to that relationship. Exit, by contrast, operates through equilibrium matching and reshapes mitigation incentives across multiple firms. Evaluations that focus narrowly on firm-level responses therefore systematically understate the impact of exit relative to engagement.

We quantify these forces in a calibration using data on firms, banks, and workers. The results show that reallocation spillovers are quantitatively important: once spillovers are taken into account, the aggregate mitigation induced by exit can be comparable to, or exceed, that generated by engagement, even when exit appears weak at the firm level. These findings imply that policies and evaluations that emphasize engagement alone may overlook a key mechanism through which values-driven stakeholders shape firm behavior.

1.1 Related Literature and Contribution

Exit vs. voice. Our paper contributes to the literature on exit and voice. Hirschman 1972 provides the canonical framework for how dissatisfied stakeholders discipline organizations, distinguishing between exit—reallocating away from an organization—and voice—attempting to change it from within. Economists have since formalized exit as a disciplinary device in

markets.[4]

More recently, exit and engagement have been contrasted most directly in models of socially responsible investment. Socially responsible funds (exit) can affect firm financing and investment decisions (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 instrument for achieving impact than voting or voice.[5]

By endogenizing exit and engagement in a general equilibrium environment with multiple types of stakeholders and heterogeneous productivity, our analysis highlights two central insights. First, we characterize which stakeholders and firms optimally choose exit versus engagement, rationalizing why large, highly productive stakeholders—such as major banks or large institutional investors—often favor exit over engagement in practice. Second, we show that evaluating exit solely at the firm level can be misleading. While exit may appear ineffective for directly targeted firms, it operates through equilibrium reallocation and generates spillovers that affect mitigation incentives economy-wide.

We further show that the cost of engagement—modeled as an endogenous subsidy provided by values-driven stakeholders to induce mitigation—depends on equilibrium matching and market competition and varies substantially across stakeholder types, beyond standard compensating differentials. This feature helps explain why, empirically, wage premia associated with values alignment can be much larger than green financing premia. By explicitly allowing for interactions among multiple stakeholder types, our framework departs from models that study a single class of stakeholders in isolation. Finally, although we do not model voting explicitly, the effectiveness of voice in our framework depends on the endogenous concentration of values-driven stakeholders within the firm. In this sense, our sorting results complement

[4]Tiebout 1956 shows how mobility ("exit") disciplines local governments. Subsequent work in political economy studies how migration and emigration function as exit mechanisms (e.g., Epple, Romer, and Sieg 2001).

[5]Relatedly, another strand of the literature emphasizes non-pecuniary motives. For example, Besley and Ghatak 2005 formalize mission preferences in organizations, showing how workers with prosocial preferences sort into mission-aligned firms. Ellingsen and Johannesson 2008 and Bénabou and Tirole 2006 study prosocial motivation, signaling, and identity, while Akerlof and Kranton 2000 incorporates identity directly into utility.

existing models by endogenizing the composition—and hence the strength—of values-driven voice.

Matching and sorting on mulit-dimensional characteristics Methodologically, our environment can be interpreted as a multi-sided matching model (with multiple groups of stakeholders) featuring multidimensional characteristics (stakeholders differ in both productivity and values) with transferable utility (we allow for price competition). Applying the optimal transport theory, the literature has established the equivalence between stable matching with transferable utility and a corresponding planner's problem (see Galichon 2016 for a comprehensive survey).

Most closely related is Boerma, Tsyvinski, and Zimin 2025, who study a multi-sided environment in which firms choose among different types of workers under a submodular output function, with workers characterized by a single dimension. In contrast, we maintain a standard supermodular production technology but allow stakeholders to differ along two dimensions—productivity and values—which interact endogenously in determining surplus and sorting.

In multidimensional matching environments, sorting patterns are generally more complex than in one-dimensional models (e.g., Sattinger 1979, Tervio 2008, Gabaix and Landier 2008), where supermodularity directly governs assortative matching.[6] In bilateral matching settings, prior work has characterized equilibrium under specific conditions with two dimensions (e.g., Dupuy and Galichon 2014, Lindenlaub 2017, Chiappori, McCann, and Pass 2016, Chiappori, Oreffice, and Quintana-Domeque 2018). Relative to this literature, we highlight how productivity and values are naturally interdependent through mitigation incentives.

Beyond bilateral matching, we contribute a tractable iterative solution method for multi-sided matching with multidimensional heterogeneity. This approach allows us to characterize how sorting and mitigation decisions interact across different types of stakeholders in general equilibrium.

[6]Some studies reduce multidimensional heterogeneity to a single index, rendering matching effectively one-dimensional.

2 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 $\mathbf{x} = (x_1, \ldots, x_{N+1})$. As discussed in Tervio (2008), the linear assumption on the arguments does not preclude different stakeholders having different skills contributing to their ability to affect output.[7]

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' utilities differ in whether they are values-driven. Value-driven stakeholders are modeled as having disutilities over the harm generated by the firm with which they have a relationship. As standard, this warm-glow preference thus creates incentives for its matching firm to mitigate. [8] Let $\theta_\ell \in \{0, 1\}$ denote type, and the utility is given by

$$u(p, e \mid \theta_\ell) = p - \theta_\ell \psi(e), \tag{1}$$

[7]Specifically, one could interpret $x_\ell = b_\ell(\hat{x}_\ell)$, which represents the effective ability of some underlying skill $\hat{x}_\ell$, where b_ℓ is an increasing transformation of the scale of measurement for a factor quality. For example, a Cobb-Douglas production function $x_0 \hat{x}_1^{\alpha} \hat{x}_2^{(1-\alpha)}$ can be nested as $x_1 = \hat{x}_1^{\alpha}$ and $x_2 = \hat{x}_2^{1-\alpha}$.

[8]Given the preferences are on firm harm within the team, the matching problem is not subject to externalities. Our framework, however, can be extended to the environment where the individual stakeholders also care about the aggregate harm (i.e., with externalities). The equilibrium outcome, however, remains the same as each agent is non-atomic and behaves as if its own choice doesn't have the aggregate harm.

where p is the transfer received, e is harm created by the hiring firm, and $\psi(\cdot)$ is increasing and convex. That is, 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 $\lambda_{N+1} = 0$. That is, firms are special in the sense that their characteristics are always one-dimensional ($\lambda_{N+1} = 0$). [9]

Mitigation. 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.

Given that we have transferrable utilities, the mitigation within any match must be chosen to maximize the total surplus. The total surplus thus yields

$$\Lambda(\mathbf{a}) = \max_{m \geq 0} \; y(\mathbf{x}) - cm - n(\boldsymbol{\theta})\psi(\sigma y(\mathbf{x}) - m). \tag{2}$$

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

$$U_\ell(a_\ell) = \max_{\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}} \Lambda(\{a_{\ell'}\}_{\ell' \in L \setminus \{\ell\}}, a_\ell) - \sum_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'}). \tag{3}$$

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 (3) with $\ell = N+1$. Figure 1 summarizes the flow of our model.

[9] We use this assumption to highlight that firms only mitigate when matching with value-driven stakeholders, and it must be profitable for them to do so. Our setup can easily be generated if firms also care about harms.

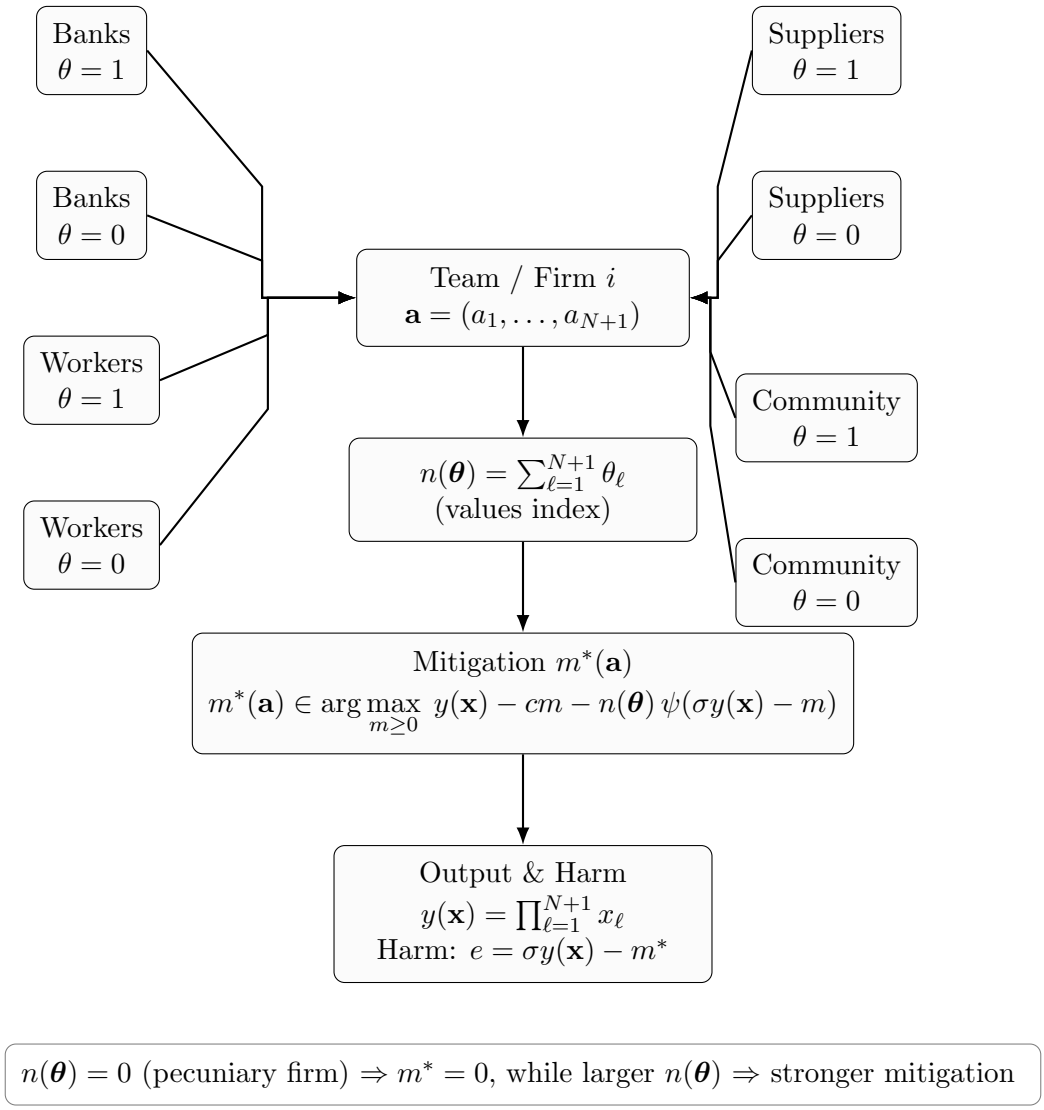

Figure 1: Extended schematic: banks and workers (left), suppliers and community (right) match with a firm, which determines the values index, mitigation, and resulting output and harm.

Competitive equilibrium. A competitive equilibrium consists of: (1) a allocation function γ represents the matching between firm and all types of stakeholders, which is define as a probability measure on the set of $(X_\ell \times \{0, 1\})^{N+1}$, (2) a mitigation policy $m^*(\mathbf{a})$ for each team, and (3) Equilibrium utilities $\{U_\ell(a_\ell)\}$ for all types $\forall \ell \in L$. These must satisfy:

1. (Optimal mitigation) The mitigation policy $m^*(\mathbf{a})$ solves Equation (2).

2. (Stable match) No matched agent can increase their utilities by forming a new team: for any $\mathbf{a} = (a_1, a_2 ... a_{N+1})$,$\Sigma_{\ell \in L} U_\ell(a_\ell) \geq \Lambda(\mathbf{a})$, where an equality obtains on the support of γ.

3. (Market clearing) The marginal of γ should equal the marginal distribution of F_ℓ.

As standard, a matching is stable if no coalition can profitably deviate to form a new team. Accordingly, for any assignment $\mathbf{a}'$ outside the support of γ, the joint surplus generated by such a deviation, $\Lambda(\mathbf{a}')$, must be weakly lower than the sum of the members' equilibrium utilities. As is well known in the matching literature with transferable utility, any stable outcome solves the planner's problem (see Boerma, Tsyvinski, and Zimin 2025). Therefore, the equilibrium maximizes aggregate surplus. Finally, given equilibrium utilities and the mitigation policy, transfers are uniquely pinned down for each stakeholder by Equation 1.

Remark on Efficiency. Our setting does not explicity feature externalities, since stakeholders internalize the harm generated by their matched firm through disutility. One could alternatively assume that stakeholders also suffer disutility from aggregate emissions. This extension would not affect the equilibrium allocation, as stakeholders are infinitesimal and take aggregate emissions as given. However, it would generally imply that the equilibrium is no longer efficient.

3 Characterization

We now analyze the surplus function and the resulting matching outcomes. Observe that a team can be characterized by two sufficient statistics: its productivity $y(\mathbf{x})$ and its stakeholder-values index $n(\boldsymbol{\theta})$. Specifically, from Equation 2, the surplus of any team can be written as

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

where

$$\Omega(y, n) = \max_{m \geq 0} \Big\{ y - cm - n\psi(\sigma y - m) \Big\}. \tag{4}$$

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* $\Omega(y(\underline{\mathbf{x}}), N) > 0$.

3.1 General Properties

Sorting on values Observe that adding a values-driven stakeholder is costly ($\Omega_n(y, n) < 0$), but the marginal cost is decreasing ($\Omega(y, n)$ is strictly convex in n). This is because mitigation is non-rival and has an economic scale.

This implies clustering: values-driven agents prefer to join each other. Let $n^*(x_\ell, \theta)$ be the index of the matching team of stakeholder (x_ℓ, θ), which consists of other types of stakeholders.

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

$$n^*(x_\ell, 1) \;\geq\; n^*(x_\ell, 0).$$

That is, conditional on the skill, values-driven stakeholders must join a team with a higher value-index than the equivalent pecuniary stakeholders. 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, and such deviation generates higher payoff as $\Omega(y, n)$ is convex in n.

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, using the envelope theorem, $\Omega_y(y, n) = 1 - n\psi'(\sigma y - m^*(y, n))\sigma$, we have

$$\frac{\partial}{\partial y}\Omega(y,n) = \begin{cases} (1-c\sigma)\,y, & n \geq 1, \\ y, & n = 0, \end{cases} \tag{5}$$

where we used the fact that, for $n \geq 1$, the interior FOC, implies $n\psi'(\sigma y - m^*(y,n)) = c$. Thus, productivity is fully valued only in purely pecuniary teams; any values-driven presence induces a constant discount. Note that our linear-cost specification preserves tractability while isolating the key In a productivity–values trade-off, the same economics remain robust as long as greater output entails greater harm. [10] [11]

Consider a stakeholder (x_ℓ, θ_ℓ) of type ℓ joining a team composed of other stakeholder types. Let (y, n) denote the productivity and stakeholder-values index of that team. The stakeholder's choice problem can be written as

$$U(x_\ell, \theta_\ell) = \max_{(y,n)} \; \Omega(x_\ell y,\, n + \theta_\ell) - T(y,n),$$

where $T(y,n)$ denotes the total utility accruing to the other members of the team.

For a pecuniary stakeholder $(x_\ell, 0)$, the marginal contribution to team surplus depends on both productivity and values composition:

$$\frac{\partial}{\partial x_\ell}\Omega(yx_\ell,\, n) = \begin{cases} (1-c\sigma)\,y, & n \geq 1, \\ y, & n = 0. \end{cases} \tag{6}$$

[10] If surplus were separable in y and n, sorting on ability and values would decouple. In our setting, higher output necessarily entails greater harm, intertwining the two dimensions. This remains robust under convex mitigation costs or harm-dependent disutility, $\Omega_y(y,n) \leq \Omega_y(y,0)$ and $\Omega_{yn}(y,n) < 0$.

[11] To be concrete, consider an alternative formulation that shuts down the dependence between productivity and mitigation, where

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

A natural interpretation of this setting is where a firm can provide costly amenities (such as office amenities) that improves workers' utilities. In this case, one can see that the surplus is separable in production y and the values-driven index n; hence, the sorting can be solved independently

This motivates the index

$$z(y,n) \equiv \begin{cases} (1-c\sigma)y, & n \geq 1, \\ y, & n = 0, \end{cases}$$

which summarizes the joint effect of productivity and values composition on the marginal return to pecuniary skill. By monotone comparative statics, higher-skill pecuniary stakeholders must therefore match with teams with higher $z(y,n)$. In particular, if a more productive pecuniary stakeholder is indifferent between a pecuniary team and a team with values-driven members, the two teams must have the same z-index, implying that the pecuniary team is less productive.

For a values-driven stakeholder $(x_\ell, 1)$, productivity is always discounted due to mitigation:

$$\frac{\partial}{\partial x_\ell}\Omega(yx_\ell,\, n+1) = (1-c\sigma)y. \tag{7}$$

As a result, a values-driven stakeholder's ranking over teams depends only on productivity y. Higher-skill values-driven stakeholders therefore match with more productive teams.

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

Unlike the standard one-dimensional benchmark, equilibrium sorting must jointly account for productivity and values. Because pecuniary stakeholders strictly prefer teams without values-driven members, equilibrium matching generally distorts productivity.

3.2 Illustrative example with $N = 2$

This section characterizes equilibrium matching. We first analyze a simple environment, where firms form relationships with two types of stakeholders (say banks and workers). We then extend the construction to general N via a sequential algorithm, and finally characterize transfers and premiums. Throughout, skills and values-driven preferences are assumed to be independently

and identically distributed.

Assumption 2. *Skills and values are independently distributed. Conditional on skill x_ℓ, a stakeholder of type ℓ is values-driven with probability λ_ℓ.*

Balanced supply: full segmentation Since we know that firms are all pecuniary, the matching between the firm and the team consisting of types ℓ and ℓ' can then be characterized by PAM on the firm's productivity $x_N + 1$ and the team's z-index. Hence, we first focus on the matching between stakeholders

When the share of values-driven stakeholders is identical across types, $\lambda_\ell = \lambda_{\ell'}$, equilibrium matching is fully segmented. By clustering (Lemma 1), values-driven stakeholders match exclusively with values-driven stakeholders, and pecuniary stakeholders match exclusively with pecuniary stakeholders. Within each segment, matching is positive assortative in skills by Lemma 2. Since firms are all pecuniary, they will match the team consisting of type ℓ and ℓ' based on their z-index.

Proposition 1 (Balanced supply)**.** *When $\lambda_\ell = \lambda_{\ell'}$, equilibrium matching is fully segmented, with positive assortative matching within each segment. And there will be a positive sorting between firms and the z-index of the team.*

Unbalanced supply: separation at the top, mixing at the bottom When $\lambda_\ell \neq \lambda_{\ell'}$, full segmentation is no feasible due to the market clearing. Suppose $\lambda_\ell > \lambda_{\ell'}$, so that type ℓ has relatively more values-driven stakeholders. Market clearing then requires some values-driven type-ℓ stakeholders to match with pecuniary type-ℓ' stakeholders. We show that such mixing occurs at the bottom of the skill distribution.

Proposition 2 (Full separation at the top)**.** *When $\lambda_\ell \geq \lambda_{\ell'}$, there exists a cutoff $\hat{x}_{\ell'}$ such that, for all $x_{\ell'} \geq \hat{x}_{\ell'}$, pecuniary (values-driven) stakeholders of type ℓ' match only with pecuniary (values-driven) stakeholders of type ℓ.*

Relative to the balanced benchmark, pecuniary type-ℓ stakeholders are scarce, so maintaining pure pecuniary teams requires sacrificing productivity. At high skill levels, pecuniary

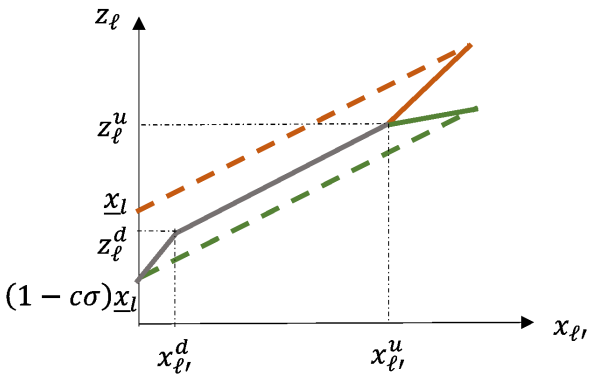

Figure 2: Dashed lines represent balanced supply, where the z-index of values-driven matches is lower due to the discount factor $(1 - c\sigma)$. The solid lines represent unbalanced supply, where full separation occurs the top. The dash values-driven (pecuniary) line represents the z-index of the matching partner under the balanced supply for values-driven (pecuniary) stakeholder of type ℓ'.

stakeholders of type ℓ' continue to prefer pure pecuniary matches, while at lower skill levels they optimally accept values-driven partners, generating mixing.

Productivity loss versus mitigation cost A pecuniary stakeholder of type ℓ' trades off matching with a less productive pecuniary partner against matching with a more productive values-driven partner while incurring mitigation costs. In Figure 2, the solid brown (green) line depicts equilibrium matching for pecuniary (values-driven) stakeholders under unbalanced supply. The dashed brown (green) line denotes the z-index of pecuniary (values-driven) matches under balanced supply.

Relative to the balanced case, the solid brown line shifts downward, reflecting reduced productivity of pecuniary matches. Nevertheless, at the top of the distribution it remains above the green line, implying that pure pecuniary matching remains optimal. Once productivity losses exceed the benefit of avoiding mitigation, pecuniary stakeholders mix with values-driven partners. As a result, mixing arises only at the bottom.

PAM between $x_{\ell'}$ and z_ℓ in the mixing region Assume that type ℓ' has excess pecuniary supply throughout the distribution, so that pecuniary type-ℓ' stakeholders are the marginal agents. Since pecuniary stakeholders rank partners by their z-index, define $\Psi_\ell(z)$ as the measure

of type-ℓ stakeholders with effective z below z:

$$\Psi_\ell(z) \equiv \int_{\underline{y}}^{z/(1-c\sigma)} g_\ell^1(\tilde{y})\, d\tilde{y} + \int_{\underline{y}}^{z} g_\ell^0(\tilde{y})\, d\tilde{y}. \tag{8}$$

Let $\phi_\ell^m(x_{\ell'})$ denote the z-index of the type-ℓ partner matched with $x_{\ell'}$. Positive assortative matching implies

$$\Psi_\ell(\phi_\ell^m(x_{\ell'})) = F_{\ell'}(x_{\ell'}), \tag{9}$$

as illustrated by the solid gray line in Figure 2.

Clustering within the mixing region Although stakeholders with a given $x_{\ell'}$ are matched with partners sharing the same z-index, the allocation must satisfy clustering (Lemma 1). Under excess pecuniary supply, scarce values-driven type-ℓ' stakeholders match exclusively with values-driven type-ℓ stakeholders, yielding teams with $n = 2$, while abundant pecuniary type-ℓ' stakeholders mix, yielding teams with $n = 1$.

Summary High-skill stakeholders remain separated, as the cost of mixing is prohibitive at the top of the distribution. Mixing occurs at the bottom, where productivity losses are small. Competition endogenously determines which stakeholders bear the cost of mitigation.

3.3 General case: sequential algorithm for $N \geq 2$

We now extend the construction to N stakeholder types using a sequential algorithm. Types are ordered by the prevalence of values-driven stakeholders, and at each stage one additional type is matched to an existing team. The matching problem at every stage mirrors the $N = 2$ case analyzed above.

Step 1: ordering of types. Label stakeholder types such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N,$$

where λ_ℓ denotes the share of values-driven stakeholders of type ℓ. Type 1 is thus the most values-driven, while later types are increasingly pecuniary.

Step 2: initial matching ($N = 2$). In the first stage, stakeholders of types 1 and 2 are matched. This reduces exactly to the $N = 2$ environment characterized in the previous section.

Step 3: iteration. Once types 1 and 2 are matched, treat each resulting pair as a fixed team. In the next stage, these teams are matched with stakeholders of type 3. Since type 3 is relatively more pecuniary, the same logic applies: full separation holds at the top of the distribution, while mixing arises at the bottom. At this stage, type-3 stakeholders are ranked by skill x_3, and teams are ranked by their effective z-index.

This procedure is repeated sequentially. At stage τ, the team formed by types $1, \ldots, \tau$, with characteristics (y_τ, n_τ), is matched with stakeholders of type $\tau + 1$, who are weakly more pecuniary. Let $x^*_{\tau+1}(y_\tau, n_\tau)$ and $\theta^*_{\tau+1}(y_\tau, n_\tau)$ denote the characteristics of the stakeholder of type $\tau + 1$ matched to a team (y_τ, n_τ). Team characteristics then evolve according to

$$y_{\tau+1} = y_\tau \cdot x^*_{\tau+1}(y_\tau, n_\tau), \qquad n_{\tau+1} = n_\tau + \theta^*_{\tau+1}(y_\tau, n_\tau). \tag{10}$$

The matching at stage τ induces a distribution of team characteristics $(y_{\tau+1}, n_{\tau+1})$ comprising types $1, \ldots, \tau + 1$. Let $g^1_{\tau+1}(y_{\tau+1})$ and $g^0_{\tau+1}(y_{\tau+1})$ denote the densities of teams with $n_{\tau+1} \geq 1$ and $n_{\tau+1} = 0$, respectively.

Final step: matching with firms. After N stages, each team contains all stakeholder types. Since firms are pecuniary by assumption, the final matching is standard: firms sort positively with teams by productivity, so that firm ability x_{N+1} is matched assortatively with the team's z-index.

Note that one key assumption for constructing PAM based on $x_{\ell'}$ and z_ℓ is that type ℓ' has relatively abundant pecuniary stakeholders so that our construction guarantees that they are the one mixing. The measure of pecuniary stakeholder conditional on $x_{\ell'}$ is constant under

i.i.d assumption, which is given by $(1-\lambda_{\ell'})$. On the other hand, the measure of pure pecuniary stakeholder conditional on z_ℓ is given by $\Pr(n_\ell = 0|z) = \frac{g^0_\ell(z)}{g^0_\ell(z)+g^1_\ell(z/(1-c\sigma))}$, which is generally not constant. Hence, the formal condition below guarantees that type ℓ' always have excess pecuniary stakeholder relatively to his matching counterparts everywhere in the distribution.

Condition 1 (excess of pecuniary stakeholder of type ℓ'): For $\ell' > \ell$, $(1-\lambda_{\ell'}) \geq \Pr(n_\ell = 0|z)$ for $\forall z \leq \phi^m_\ell(\hat{x}_{\ell'})$.

Proposition 3 (Sequential construction of equilibrium)**.** *Suppose Condition 1 holds at every stage τ. Then the equilibrium matching outcome can be constructed sequentially as follows. At each stage τ, the matching between teams (y_τ, n_τ) and stakeholders $(x_{\tau+1}, \theta_{\tau+1})$ is characterized by a cutoff $\hat{x}_{\tau+1}$ such that:*

1. Full separation at the top: *pecuniary (values-driven) stakeholders above the cutoff match exclusively with teams with $n_\tau = 0$ ($n_\tau \geq 1$).*

2. Mixing at the bottom: *below the cutoff, matching is characterized by positive assortative matching between stakeholder skill $x_{\tau+1}$ and the team z-index $z_\tau = z(y_\tau, n_\tau)$. Conditional on $(z_\tau, x_{\tau+1})$, values-driven stakeholders are matched first with teams with higher values-driven indices.*

Team characteristics evolve according to (10).

Evolution of values-driven teams. The sequential structure implies two implications. First, once a team becomes purely pecuniary ($n_\tau = 0$), it remains pecuniary in all subsequent stages, since later stakeholder types are weakly more pecuniary. As a result, the total mass of pecuniary teams is $(1-\lambda_1)$, the mass of pecuniary type-1 stakeholders. Second, in later stages values-driven stakeholders are increasingly scarce. Consequently, only teams with higher values-driven indices continue to match with values-driven stakeholders, so that values-driven intensity becomes progressively concentrated among a shrinking subset of teams.

3.4 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 3 implies that

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

where $z^*(x_\ell, \theta)$ denote the z index of the optimal team of stakeholder (x_ℓ, θ), which consists all types of stakeholders excluding himself. Thus, the marginal gain from ability equals the effective productivity (z-index) of the matching team. It also implies that, unless the team is purely pecuniary, all stakeholders' marginal contributions are discounted by $(1 - c\sigma)$.

Hence, given the matching $z^*(\tilde{x}_\ell, \theta)$, the utility for stakeholder (x_ℓ, θ) relative to the lowest ability type, is thus uniquely pinned down

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

Compensation For pecuniary agents $(x_\ell, 0)$, equilibrium utility coincides with the transfer they receive: $p_\ell(x_\ell, 0) = U_\ell(x_\ell, 0)$. 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 $(x_\ell, 1)$ 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. The compensation for $(x_\ell, 1)$ who is matched with the firm with index-value n is thus given by

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

That is, more harmful teams must pay higher wages to attract values-driven stakeholders.

Ranking premium, beyond compensating differential We now look at the difference in fees between two stakeholders with the same skills but different preferences. Recall that the logic behind the standard compensating differential is the following: when a firm is indifferent

between hiring $(x_\ell, 0)$ and $(x_\ell, 1)$, then they must be indifferent between paying for mitigation vs. hiring pecuniary stakeholders. Thus, pecuniary stakeholders should receive higher compensation conditional on skill, and the difference in fees can be explained by the mitigation costs.

In an environment where stakeholders have different abilities, firms will not generally be indifferent. In particular, for pecuniary stakeholders that are relatively scarce (i.e., type 1 stakeholder), our equilibrium construction implies that they will stay in a pure pecuniary team in equilibrium, and thus, they will be able to match with more productive firm, and thus earn higher wage relatively to his counterpart. On the other hand, for pecuniary stakeholders that are relatively abundant (i.e., type $\ell \geq 2$), we know that certain pecuniary and values-driven stakeholders of the same type may end up matching with teams of the same z-index. In this case, they will be matched with firms with the same productivity, and pecuniary stakeholders no longer capture additional rents: they are paid the same as otherwise identical values-driven stakeholders, apart from the compensating differential.

The role of matching outcomes on the fee can be formally seen below,

$$\begin{aligned} p_\ell(x_\ell, 0) - p_\ell(x_\ell, 1|n) &= U_\ell(x_\ell, 0) - U_\ell(x_\ell, 1) - \psi(\xi_n^*) \\ &= \int_{\underline{x}_\ell}^{x_\ell} \{z^*(\tilde{x}_\ell, 0) - z^*(\tilde{x}_\ell, 1)\}\, d\tilde{x}_\ell + \{U_\ell(\underline{x}_\ell, 0) - U_\ell(\underline{x}_\ell, 1)\} - \psi(\xi_n^*), \end{aligned}$$

where $\{U_\ell(\underline{x}_\ell, 0) - U_\ell(\underline{x}_\ell, 1)\}$ is the constant term with respect to x_ℓ.

Recall that in the mixing region $z^*(\tilde{x}_\ell, 0) = z^*(\tilde{x}_\ell, 1)$ for $x_\ell < \hat{x}_\ell$ and for $\ell \geq 2$. Hence, the first term is zero. Otherwise, $z^*(\tilde{x}_\ell, 0) > z^*(\tilde{x}_\ell, 1)$. Whenever this happens, the pecuniary stakeholders earn additional rents as they can match with a more productive team.

Proposition 4 (Ranking Premiums for pecuniary stakeholders)**.** *Pecuniary stakeholders of type* 1 *and for the relatively talented stakeholders* $x_\ell > \hat{x}_\ell \ \forall \ell \geq 2$*, earn a positive ranking premium, where* $z^*(x_\ell, 0) > z^*(x_\ell, 1)$*. By contrast, pecuniary stakeholders in the mixing region* $(x_\ell \leq \hat{x}_\ell \ \forall \ell \geq 2)$ *earn no additional premium beyond the compensating differential.*

4 Exit vs. Engagement in General Equilibrium

We now use the equilibrium characterization to study how an exogenous increase in the share of values-driven stakeholders affects firm–stakeholder relationships and firm harm. Formally, consider a shock that increases the measure of values-driven stakeholders of type ℓ (i.e., λ_ℓ). The shock allows us to distinguish whether affected stakeholders optimally remain with their current firms (*engagement*) or reallocate to new firms (*exit*).

In contrast to the existing single-firm approach, our framework explicitly incorporates equilibrium competition. Whether a stakeholder engages or exits depends not only on the cost of engagement, but also on the quality of firms available upon exit. We show that engagement dominates at the bottom of the ability distribution, while exit occurs at the top: treated stakeholders of type ℓ optimally engage (exit) when their ability is sufficiently low (high).

Proposition 5 (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 such that: (i) treated stakeholders below the cutoff remain with a firm of the same size (engagement); (ii) treated stakeholders above the cutoff exit to smaller firms with higher stakeholder-values indices.*

Illustration: the $N = 2$ case. To illustrate the mechanism, consider $N = 2$. When the supply of pecuniary type-2 stakeholders falls, the separation threshold in Figure 2 shifts downward. Intuitively, reduced competition for pecuniary type-1 stakeholders makes pecuniary matching more attractive, delaying the onset of mixing. Let $\hat{x}_{\ell,0}$ and $\hat{x}_{\ell,1}$ denote the separation cutoffs before and after the shock. Then $\hat{x}_{\ell,1} < \hat{x}_{\ell,0}$.

Engagement at the bottom. Consider a treated stakeholder with ability $x_\ell < \hat{x}_{\ell,1}$. By construction, both before and after the shock, stakeholders $(x_\ell, 0)$ and $(x_\ell, 1)$ are matched with teams of the same z-index.[12] Therefore, when a low-ability stakeholder becomes values-driven,

[12] In the mixing region, the sorting function $\phi^m_{\ell-1}(x_\ell)$ that solves (9) is independent of λ_ℓ.

the ranking of his equilibrium match does not change: $z^*(x_\ell, 1) = z^*(x_\ell, 0)$. He can thus remain with a firm of the same size, which we interpret as engagement.[13]

Exit at the top. At the top of the distribution, full separation implies that the stakeholder-values index is binary, $n \in \{0, N\}$. Treated stakeholders previously matched with large firms must now reallocate to firms with higher values-driven indices. Moreover, once a stakeholder becomes values-driven, he no longer enjoys a ranking premium. As a result, treated high-ability stakeholders optimally exit to smaller firms after the shock.

4.1 Impact on Harm

We now study how preference shocks affect firm harm. Recall that firm harm depends on productivity y and the stakeholder-values index n:

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

When $n \geq 1$, harm depends only on n and not on productivity. Hence, the first-order impact of preference shocks operates through changes in the distribution of the stakeholder-values index across firms. For purely pecuniary firms, harm also depends on productivity, but this channel is quantitatively minor. We therefore focus on the distribution of n.

4.1.1 Aggregate Impact

Because values-driven stakeholders cluster, the distribution of firms by stakeholder-values index is mechanically determined by $(\lambda_1, \ldots, \lambda_N)$. For example, when $N = 2$, the measure of firms with two values-driven stakeholders equals λ_2, while the measure with exactly one equals $\lambda_1 - \lambda_2$. An increase in λ_2 by δ raises the measure of fully values-driven firms by δ and reduces the

[13] For pecuniary stakeholders $(x_\ell, 0)$, the equilibrium z-index is uniquely pinned down, although they may be indifferent across teams with different n. Conditional on z, all such allocations are payoff equivalent. We therefore define engagement as remaining with a firm of the same size.

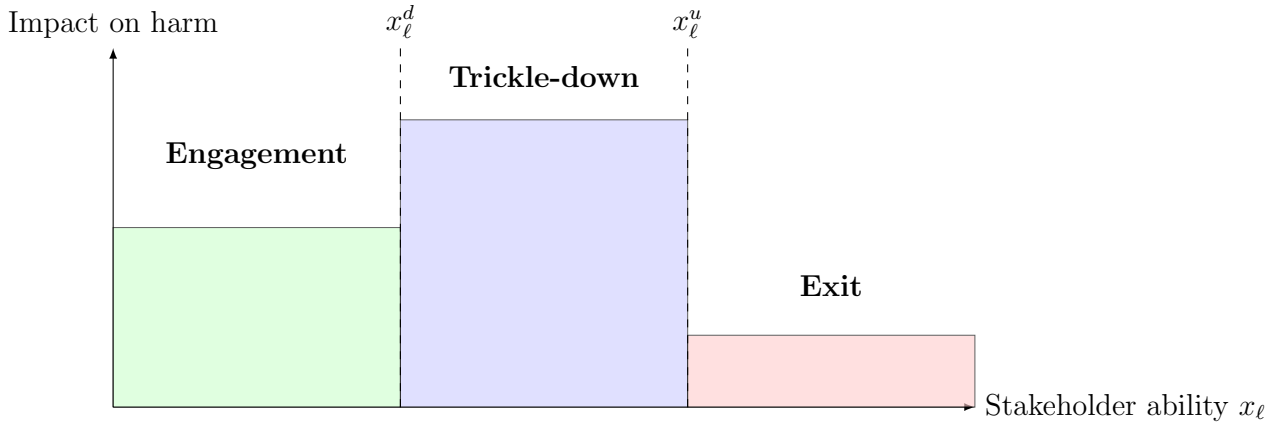

Figure 3: Schematic of micro-level impact of a values-driven preference shock. At the bottom ($x_\ell \leq x_\ell^d$), treated stakeholders stay and reduce harm directly (engagement). In the middle ($x_\ell^d < x_\ell < x_\ell^u$), displacement of untreated stakeholders generates additional harm reduction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

measure of partially values-driven firms by δ, leaving purely pecuniary firms unchanged. This logic generalizes to any N.

Proposition 6 (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 $1 - \lambda_1$. Conditional on the ordering of λ_ℓ remaining unchanged, 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.*

4.1.2 Micro-Level Impact

We next study the impact at the stakeholder level. Consider a mass δ of treated type-ℓ stakeholders who switch from pecuniary to values-driven preferences. Their effect depends on whether they engage or exit, as characterized in Proposition 5.

When a low-ability stakeholder becomes values-driven, his equilibrium match does not change. He therefore remains with the same firm, which now has one additional values-driven member. This lowers firm harm, which we refer to as the *engagement effect*. By contrast, treated high-ability stakeholders must exit to firms with higher values-driven indices. This may have no direct effect if the receiving firm already has values-driven stakeholders, but it displaces untreated values-driven stakeholders downward in the distribution.

Trickle-down effect in the middle. This displacement generates additional harm reduction in the middle of the distribution. Treated high-ability stakeholders crowd out lower-ability untreated ones, who then join smaller firms and make those firms more values-driven. As a result, harm reduction in the middle can exceed the measure of treated stakeholders. In particular, firms with $n = 1$ in the region $x_\ell \in [\hat{x}_{\ell,0}, \hat{x}_{\ell,1}]$ before the shock must have $n = 0$ or $n = 2$ afterward, since they lie above the new cutoff $\hat{x}_{\ell,1}$.

Proposition 7 (Trickle-down effect)**.** *(i) At the bottom ($x_\ell \leq \hat{x}_{\ell,1}$), harm reduction equals the measure of treated stakeholders. (ii) In the middle region ($x_\ell \in [\hat{x}_{\ell,0}, \hat{x}_{\ell,1}]$), harm reduction exceeds the measure of treated stakeholders due to displacement of untreated ones. (iii) At the top ($x_\ell \geq x^u_\ell$), some treated stakeholders have no direct effect on harm.*

Aggregate harm reduction equals $\delta(\xi^*_\ell - \xi^*_{\ell-1})$ (Proposition 6). Since engagement at the bottom accounts exactly for $\delta(\xi^*_\ell - \xi^*_{\ell-1})$, and not all treated stakeholders at the top have direct effects, harm reduction in the middle must exceed this amount. Thus, the aggregate impact is realized disproportionately in the middle of the distribution, even though the preference shock is uniformly distributed.

4.2 Spillover Effects on Payoffs

We finally examine how preference shocks affect equilibrium payoffs. Although only type-ℓ stakeholders are directly treated, the induced reallocation alters competition and matching economy-wide.

An increase in λ_ℓ intensifies competition among values-driven type-ℓ stakeholders, lowering the z-index of their equilibrium matches at the top of the distribution (the green line in Figure 2 shifts downward). Conversely, pecuniary type-ℓ stakeholders become scarce and match with higher-ranked teams. Thus, value-driven (pecuniary) type-ℓ stakeholders at the top are worse (better) off.

For stakeholders of other types $\ell' \neq \ell$, the effect is reversed. Due to complementarities in production, improvements in the skill distribution of values-driven teams raise productivity

and benefit values-driven stakeholders of other types. Pecuniary stakeholders of other types, however, face less productive matches.

Proposition 8 (Spillover effects on payoffs)**.** *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 $x^e_{\ell'}$ for each ℓ' such that: (i) for type ℓ, value-driven (pecuniary) stakeholders experience lower (higher) payoffs if $x_\ell \geq x^e_\ell$; (ii) for any $\ell' \neq \ell$, value-driven (pecuniary) stakeholders experience higher (lower) payoffs if $x_{\ell'} \geq x^e_{\ell'}$. Payoffs are unchanged for all low-ability stakeholders ($x_{\ell'} \leq x^e_{\ell'}$).*

Engagement does not generate payoff spillovers, since matching at the bottom remains unchanged. However, because engagement reduces firm harm, it implies a redistribution within the firm: monetary payments to other values-driven stakeholders must decline.

5 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).[14]

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

[14] 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 $\kappa = 8 \times 10^{-12}$, $\rho = 1$.

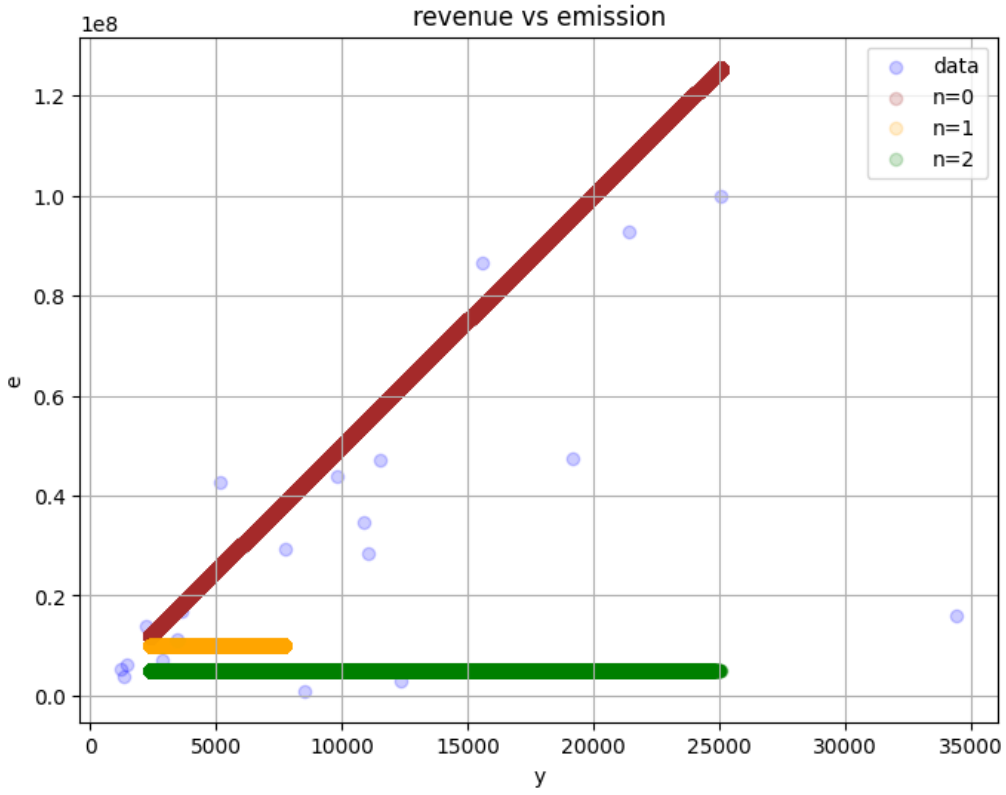

Figure 4: Calibrated production–emission relation.

all treated banks engaged their original firms, 8/15 of green banks (53%) would have impact. 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.

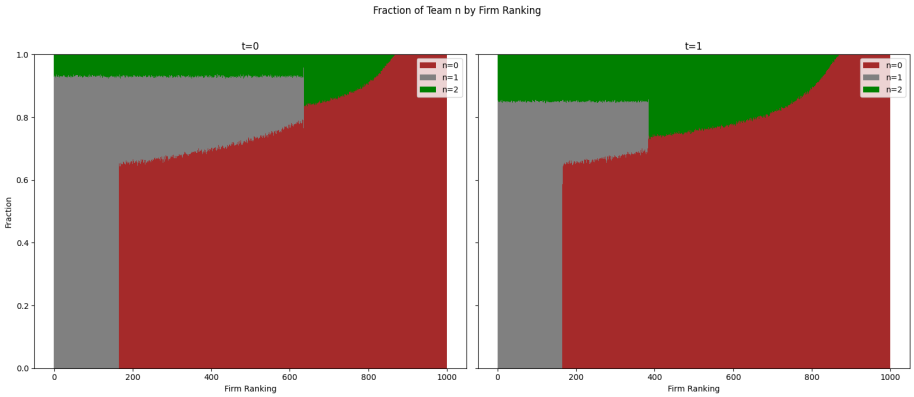

Figure 5: Firm matching outcomes before vs. after the shock.

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

6 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.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin (2025). "Sorting with Teams". In: *Journal of Political Economy* 133.2, pp. 421–454.

Bonnefon, Jean-François et al. (2025). "The moral preferences of investors: Experimental evidence". In: *Journal of Financial Economics* 163, p. 103955.

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.

Galichon, Alfred (2016). *Optimal transport methods in economics*. Princeton University Press.

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*.

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.

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.

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 Proof of Lemma 1

Proof. Observe that

$$\Omega(y,n) = \max_{e \leq \sigma y} \{y - n\psi(e) - c(\sigma y - e)\}$$

is decreasing and convex in n. This follows because

$$f(n, e \mid y) \equiv y - n\psi(e) - c(\sigma y - e)$$

is linear in n, and therefore $\Omega(y,n) = \max_e f(n, e \mid y)$ is (strictly) convex in n. This argument holds for a general cleaning cost function.

With the specified linear cost, $\Omega(y,n)$ can be written as

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

where

$$\chi(y,n) \equiv \max_{e \geq c\sigma y} \{ce - n\psi(e)\} = c\xi_n^* - n\psi(\xi_n^*) \quad \text{for } n \geq 1,$$

and $\chi(y,0) = c\sigma y$. This representation implies that $\chi_n < 0$ and that $\chi(y,n)$ is convex in n (i.e., $\chi_{nn} > 0$). The key properties of $\chi(y,n)$ are summarized in the following lemma.[15]

Lemma 3. *For all n, $\chi(y,n) - \chi(y,n+1)$ is decreasing in n. Moreover, $\chi(y,n) - \chi(y,n+1)$ is independent of y for $n \geq 1$, and depends positively on y only when $n = 0$.*

We now prove Lemma 1 by contradiction. Suppose that a values-driven agent $(x_i, 1)$ is matched to a team with a lower values-driven index than that of an otherwise identical pecuniary agent $(x_j, 0)$, so that $n_{-i} < n_{-j}$. We show that a profitable deviation exists in which the two agents switch teams.

[15] Assumption 1 implies that it is optimal for any team to mitigate whenever it contains at least one values-driven stakeholder. More generally, similar properties hold as long as an interior solution exists for all $n \geq \hat{n}$. In that case, $\chi(y,n) = c\sigma y$ for all $n < \hat{n}$.

Intuitively, since both agents have the same ability, swapping their teams does not affect team productivity. However, because $\chi(y,n)$ is convex in n, the switch makes the distribution of the values-driven index more extreme, thereby increasing total surplus.

Formally, letting $x_i = x_j = x$, the change in total surplus from switching is

$$\begin{aligned}&\Big\{\Omega(y_{-i}x_j, n_{-i}) + \Omega(y_{-j}x_i, n_{-j}+1)\Big\} - \Big\{\Omega(y_{-i}x_i, n_{-i}+1) + \Omega(y_{-j}x_j, n_{-j})\Big\}\\ =&\Big\{\chi(y_{-i}x, n_{-i}) + \chi(y_{-j}x, n_{-j}+1)\Big\} - \Big\{\chi(y_{-i}x, n_{-i}+1) + \chi(y_{-j}x, n_{-j})\Big\}.\end{aligned}$$

When $n_{-j} > n_{-i} > 0$, Lemma 3 implies that

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

so the surplus gain is strictly positive.

It remains to consider the case in which $n_{-i} = 0$ and $n_{-j} > 0$. In this case,

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

where the inequality follows from Lemma 3 and the fact that $\chi(y,0) - \chi(y,1)$ is increasing in y. Hence, the deviation is again profitable, contradicting stability.

Therefore, the claimed clustering result must hold in equilibrium. □

A.2 Proof for Lemma 2

Proof. A stakeholder's problem can be rewritten as choosing his team optimally with characteristics (y,n) 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 can be expressed as

$$U_\ell(x_\ell, \theta_\ell) = \max_{(y,n)} \Omega(yx_\ell, n + \theta_\ell) - \Sigma_{\ell' \in L \setminus \{\ell\}} U(a_{\ell'}).$$

Since Equation 7 implies complementarity between green agent $(x_\ell, 1)$ and y, hence, by the monotonic comparative statics, a green agent with higher ability must choose a team with a higher productivity than a green agent with lower ability. Similarly, Equation 6 implies complementarity between brown agent $(x_\ell, 0)$ and $z(y, n)$; hence, a more skilled brown agent must choose a team with a higher z-index. □

A.3 Proof for Unbalanced Supply

Characterization of Cutoff. Let $G_\ell(y_\ell, n_\ell)$ denote the cumulative distribution for the team with characteristics (y_ℓ, n_ℓ) at period ℓ, and $G^0_\ell(y)$ and $G^1_\ell(y)$ denote the measure of team with index $n_\ell = 0$ and $n_\ell \geq 1$ with productivity no larger than y, respectively. Let $\phi^\theta_s(x_{\ell'})$ denote the productivity of stakeholder $(x_{\ell'}, \theta)$ in the full separation region, where

$$G^0_\ell(\bar{y}_\ell) - G^0_\ell(\phi^0_s(x_{\ell'})) = (1 - \lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x_{\ell'})),$$

and

$$G^1_\ell(\bar{y}_\ell) - G^1_\ell(\phi^1_s(x_{\ell'})) = \lambda_{\ell'}(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x_{\ell'})). \tag{11}$$

The cutoff can be determined in two ways. The first case is when all pure pecuniary teams are allocated to stakeholder $(x_{\ell'}, 0)$ above the cutoff. Thus,

$$G^0_\ell(\bar{y}_\ell) - G^0_\ell(y^0_{\ell,L}) = (1 - \lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x^0_{\ell',L})), \tag{12}$$

where $y^0_{\ell,L}$ denote the lowest productivity for team with index $n = 0$. If $y^0_{\ell,L} > (1 - c\sigma)\phi^1_s(x^0_{\ell,L})$,

then $\hat{x}_{\ell'} = x^0_{\ell,L}$. If not, then there exists an interior solution where $\hat{x}_{\ell'}$ is given by

$$\phi^0_s(\hat{x}_{\ell'}) = (1 - c\sigma)\phi^1_s(\hat{x}_{\ell'}). \tag{13}$$

Given any $G_\ell(y_\ell, n_\ell)$, the interior solution exists whenever $0 < \lambda_{\ell'} < \hat{\lambda}_{\ell'}$. This is because that when type ℓ' has more brown stakeholders, the cutoff $x^0_{\ell',L}$ must increase, according to Equation 12. Moreover, Equation 11 also implies that a lower $\lambda_{\ell'}$ results in a higher $\phi^1_s(x_{\ell'})$, as a value-driven stakeholder can hire more productive team.

Lemma 4. *Under condition* $\lambda_{\ell'} \geq \frac{dG^0_\ell(y)}{dG^0_\ell(y)+dG^1_\ell(\frac{y}{1-c\sigma})}$, *the stable matching between team* (y_ℓ, n_ℓ) *and stakeholder* $(x_{\ell'}, \theta_{\ell'})$ *at any period* ℓ *can be constructed by a cutoff* $\hat{x}_{\ell'}$,*where* $\hat{x}_{\ell'}$,*where (1) above the cutoff off, there is full separation, where* $n_\ell = 0$ *(*$n_\ell \neq 0$*) is only matched with stakeholder with* $\theta_{\ell'} = 0$ *(*$\theta_\ell = 1$*),and conditional on* $\theta_{\ell'}$,*the matching is characterized by PAM between productivity* y_ℓ *and* $x_{\ell'}$*; (2) below the cutoff, the matching can be characterized by PAM between* z_ℓ *and* $x_{\ell'}$,*and conditional on* $(z_\ell, x_{\ell'})$*, the value-driven stakeholder* $(x_{\ell'}, 1)$ *is matched with the team with higher* n_ℓ*. For relative small* $\lambda_{\ell'} < \hat{\lambda}_{\ell'}$ *the cutoff* $\hat{x}_{\ell'}$ *is given Equation 13, otherwise, the cutoff solves Equation 12.*

Proof. The constructed matching guarantees that Lemma 1 and 2 are satisfied. Hence, fixing the preference, there is no profitable deviation for (x_ℓ, θ_ℓ) to switch with another stakeholder with the same preference due to complementarity. What is left to show is no profitable deviation to switch across preference.

Observe that our equilibrium construction implies that $z^*_\ell(x, 0) \geq z^*_\ell(x, 1)$. That is, the matching team of the pecuniary stakeholder $(x, 0)$ must have higher $z-$index than the one of $(x, 1)$, even though their productivity might be lower. This allows us to show that there is no profitable matching due to complementariy. Specifically, we first consider the case of switching between $(x_i, 1)$ and $(x_j, 0)$ of type ℓ',where $x_i \leq x_j$. That is, a value-driven stakeholder switches with a pecuniary stakeholder with higher ability. Let $y^*_i \equiv y^*_\ell(x_i, \theta_i)$, $n^*_i \equiv n^*_\ell(x_i, \theta_i)$, and $n^*_i \equiv z^*_\ell(x_i, \theta_i)$ denote the characteristics of the matching team of stakeholder i of type ℓ'. The

total surplus after switches thus yields

$$\begin{aligned}
&\Omega(y_i^* x_j, n_i^*) + \Omega(y_j^* x_i, n_j^* + 1) \\
\leq &\{z_i^* x_j + \chi(y_i^* x_i, n_i^*)\} + \left\{z_j^* x_i + \chi(y_j^* x_i, n_j^* + 1)\right\} \\
\leq &\left\{z_j^* x_j + z_i^* x_i\right\} + \chi(y_i^* x_i, n_i^*) + \chi(y_j^* x_i, n_j^* + 1) \\
< &\left\{z_j^* x_j + z_i^* x_i\right\} + \chi(y_i^* x_i, n_i^* + 1) + \chi(y_j^* x_i, n_j^*) \\
= &\Omega(y_i^* x_i, n_i^* + 1) + \Omega(y_j^* x_j, n_j^*).
\end{aligned}$$

The first equality uses the fact $\Omega(y, n+1) = (1 - c\sigma)y + \chi(y, n+1)$ and $z_j^* \geq (1 - c\sigma)y_j^*$, where the inequity holds if $n_j^* \geq 1$. The second inequality uses the complementarity of zx, as $z_j^* = z_\ell^*(x_j, 0) \geq z_\ell^*(x_j, 1) \geq z_\ell^*(x_i, 1) = z_i^*$ if $x_j \geq x_i$. The last inequality uses the fact that $\chi(y, n)$ is convex in n, and value-driven agent i must be matched with team with higher index (i.e., $n_{-i}^* \geq n_{-j}^*$). Thus, according to Lemma 3, for any $n_{-i}^* \geq n_{-j}^*$, $\chi(y, n_{-i}^*) - \chi(y, n_{-i}^* + 1) \leq \chi(y, n_{-j}^*) - \chi(y, n_{-j}^* + 1)$.

Next, consider the case where $x_i \geq x_j$. Note that our equilibrium construction guarantees that $y_i^* = y_\ell^*(x_i, 1) \geq y_\ell^*(x_j, 1) \geq y_\ell^*(x_j, 0) = y_\ell^*$, when $x_i \geq x_j$. That is, a value-drive stakeholder must get a more productive match than its counterpart of a pecuniary stakeholder, as $\lambda_{\ell'} > \lambda_\ell$. Hence, due to the complementary of yx, there is no gain in productivity when matching across markets. Specifically, such deviation must yield lower total surplus, as

$$\begin{aligned}
&\Omega(y_i^* x_j, n_i^*) + \Omega(y_j^* x_i, n_j^* + 1) \\
= &(1 - c\sigma)\left(y_i^* x_j + y_j^* x_i\right) + \chi(y_i^* x_i, n_i^*) + \chi(y_j^* x_i, n_j^* + 1) \\
\leq &(1 - c\sigma)\left(y_i^* x_j + y_j^* x_i\right) + \chi(y_i^* x_i, n_i^* + 1) + \chi(y_j^* x_i, n_j^*) \\
= &\Omega(y_i^* x_i, n_i^* + 1) + \Omega(y_j^* x_j, n_j^*),
\end{aligned}$$

The first equality uses the fact that $\chi(y, n)$ is independent of y for $n \geq 1$. The second inequality uses the complementarity of yx,as $y_i^* \geq y_j^*$ when $x_i \geq x_j$. The last ineqaultiy uses again that

$\chi(y, n)$ is convex in n, which captures the cost of mixing preference. □

A.4 Proof for Proposition Balanced Supply and Unbalanced Supply

Proof. In the case that $N = 2$, then $G^0_\ell(y) = (1 - \lambda_\ell)F_\ell(y)$ and $G^1_\ell(y) = \lambda_\ell F_\ell(y)$. Hence, Proposition follows directly from Lemma 4. The balanced supply is a special case where Equation 12 can be written as

$$(1 - \lambda_\ell)(F_\ell(\bar{x}_\ell) - F_\ell(\underline{x_\ell})) = (1 - \lambda_\ell) = (1 - \lambda_{\ell'})(F_{\ell'}(\bar{x}_{\ell'}) - F_{\ell'}(x^0_{\ell',L})),$$

and thus $x^0_{\ell',L} = \underline{x_{\ell'}}$ when $\lambda_\ell = \lambda_{\ell'}$. Hence, full separation between stakeholders must hold. □

A.5 Proof for Proposition 3

Proof. We know show that Lemma 4 can applied to any N under the sequential ordering. Observe that the sequential ordering implies the following properties: (1) for any brown team at period τ, their team remains brown 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 browner. (2) Given any team (y_τ, n_τ),where $n_\tau \geq 1$, we have $n_{\tau+1} = n_\tau + 1$ if it's in the full separation region (i.e., $(1 - c\sigma)y_\tau \geq \phi^1_s(\hat{x}_{\ell'})$). If it's in the mixing region, then conditional on $(1 - c\sigma)y_\tau \leq \phi^1_s(\hat{x}_{\ell'})$, only the team with relatively high value-index can be matched with a value-driven stakeholder and thus $n_{\tau+1} = n_\tau + 1$. Otherwise, $n_{\tau+1} = n_\tau$.

As a result, for all teams, the evolution of their z index can be expressed as $z_{\tau+1} = z_\tau(y_\tau, n_\tau)x^*_{\tau+1}(y_\tau, n_\tau)$,[16] and $n_{\tau+1} = n_\tau + \theta^*_{\tau+1}(y_\tau, n_\tau)$. Let $X^*_\tau(y_\tau, n_\tau)$ and $N^*_\tau(y_\tau, n_\tau)$ represent the optimal productivity and green index chosen by the team S_τ from period τ to period N.

[16]Specifically, for $n \geq 1$, we have $z_{\tau+1}(y, n) = (1 - c\sigma)yx^*_{\tau+1}(y, n) = z_\tau(y, n)x^*_{\tau+1}(y_\tau, n_\tau)\, n_\tau \geq 1$ and for $n = 0$, we have $z_{\tau+1}(y, 0) = yx^*_{\tau+1}(y, 0) = z_\tau(y, 0)x^*_{\tau+1}(y, 0)$. Importantly, this is not true if property (1) does not hold. This is because that, if a brown team receives a green stakeholder a period τ', then $z_{\tau+1} = z_\tau(1 - c\sigma)x^*_{\tau+1}(S_\tau)$.

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

Given that $X^*_{\tau+1}\left(y_{\tau+1,} n_{\tau+1}\right)$ is monotonic in $z_{\tau+1}(y_{\tau+1,} n_{\tau+1})$ and $N^*_{\tau+1}\left(y_{\tau+1,} n_{\tau+1}\right)$ is monotonic in $n_{\tau+1}$, the surplus function that takes into account all matching after period τ can be rewritten as

$$\Omega\left\{\left(y_\tau, n_\tau\right), \left(x_{\tau+1}, \theta_{\tau+1}\right)\right\} = \Gamma_y(z_\tau\left(y_\tau, n_\tau\right) x_{\tau+1}) + \Gamma_n\left(\left(n_\tau + \theta_{\tau+1}\right)\right),$$

where $\Gamma'_y(z_{\tau+1}) > 0$ and $\Gamma''_y(z_{\tau+1}) > 0$, and $\Gamma'_n(n_{\tau+1}) < 0$ and $\Gamma''_n(n_{\tau+1}) < 0$. Hence, for each period τ, the matching outcome is stable as long as it maximizes the product of $z_{\tau+1} = z_\tau x_{\tau+1}$ and the dispersion of $n_{\tau+1}$ at period τ. Thus, the same derivation for Lemma 4 continues to hold for each period τ. □

A.6 Proof for Proposition 4

Proof. Since our equilibrium implies that if $x_\ell < \hat{x}_\ell$, then the pecuniary stakeholder will be mixing between pecuniary and values-driven stakeholders with the team with same $z_{\ell-1}$ at period ℓ. As a result, they will have the same z_τ after the matches. Specifically, for $n \geq 1$,

$$z_{\tau+1}(y, n) = (1 - c\sigma) y x^*_{\tau+1}(y_{,} n) = z_\tau(y, n) x^*_{\tau+1}(z_\tau(y, n))\ n_\tau \geq 1,$$

where the second equality uses the fact that in the mixing region, there is one-to-one mapping to $x_{\tau+1}$ and z_τ. For $n = 0$, using the fact that the later stakeholder added to the team must have $\theta_\ell = 0$,we thus have

$$z_{\tau+1}(y, 0) = y x^*_{\tau+1}(y_{,} 0) = z_\tau(y, 0) x^*_{\tau+1}(z_\tau(y, 0)).$$

Hence, in both cases, the evolution of $z_{\tau+1}$ only depends on z_τ as long as x_ℓ is at the mixing region. Hence, $z^*(x_\ell, 0) = z^*(x_\ell, 1)$. □

A.7 Proof for Proposition 5

Proof. First of all, observe that for stakeholders that are relatively small, where $\forall x_\ell < \hat{x}_{\ell,1}$, the z-index of their equilibrium matching team must remain the same given the preference shocks on λ_ℓ. This is because these stakeholders are always at the mixing region before and after the shocks. Specifically, (1) at period ℓ, the matching is based on $(z_{\ell-1}, x_\ell)$,where $z_{\ell-1}$ remains the same, and x_ℓ is independent of θ_ℓ. Thus, their z-index after the match $z_\ell = z^*_{\ell-1}(x_\ell, 1)x_\ell = z^*_{\ell-1}(x_\ell, 0)x_\ell$ remains the same, and (2) for any period $\ell' \geq \ell$, the matching is based on $z_{\ell'}$ and $x_{\ell'+1}$, which is again remains the same. Hence, for relatively small stakeholders, the z-index of their matching team remains the same before and after the shock.

On the other hand, $\forall x_\ell > \hat{x}_{\ell,1}$,the equilibrium z-index $z^*(x_\ell, \theta)$ for stakeholder (x_ℓ, θ) must decrease form the value-driven stakeholder $(x_\ell, 1)$ (increase for the remaining pecuniary stakeholders $(x_\ell, 0)$), as there are more (less) competition. Hence, the measure of stakeholders $\tilde{x}_\ell$ that have a higher z-ranking than x_ℓ after the shock must increase, which means that treated $(x_\ell, 1)$ must now match with a smaller firm. □

A.8 Proof for Lemma A.8

We first establish the following lemma that shows that the share of firms in the mid-range that is matched with a value-driven stakeholder of type ℓ is strictly larger than λ_ℓ. For any $x_{N+1} < x^*_{N+1}(\hat{x}_{\ell,t})$, the share of firm below x_{N+1} that are matched with a value-driven stakeholder of type ℓ is λ_ℓ. For any $x_{N+1} \in (x^*_{N+1}(\hat{x}_{\ell,t}), \bar{x}_{N+1})$, the share of firms between $[x^*_{N+1}(\hat{x}_{\ell,t}), x_{N+1}]$ that is matched with a value-driven stakeholder is strictly larger than λ_ℓ.

Proof. Let $\psi^\theta_\ell(x_{N+1})$ denote the ability of firm-x_{N+1}'s matching stakeholder of type ℓ with

preference θ. The market clearing implies that

$$\int_{x_L}^{x} dF_{N+1}(\tilde{x}) = \int_{\underline{x}_\ell}^{\psi_\ell^1(x)} \lambda_\ell dF_\ell(\tilde{x}) + \int_{\underline{x}_\ell}^{\psi_\ell^0(x)} (1-\lambda_\ell) dF_\ell(\tilde{x}).$$

Let $P_\ell^1(x) \equiv \frac{\int_{\underline{x}_\ell}^{\psi_\ell^1(x)} \lambda_\ell dF_\ell(\tilde{x})}{\int_{x_L}^{x} dF_{N+1}(\tilde{x})}$ represent the proportion of green firms below x that are matched with value-driven stakeholders ℓ. In general, $\psi_\ell^1(x_{N+1}) \geq \psi_\ell^0(x_{N+1})$, that is, if he matches with $(x_\ell, 0)$ and $(x'_\ell, 1)$, it must be the case that the value-driven stakeholder has (weakly) higher ability. Hence, whenever the firm is indifferent between stakeholders with different preference, we thus have

$$P_\ell^g(x) = \frac{\lambda_\ell \left\{ \int_{\underline{x}_\ell}^{\psi_\ell^1(x)} dF_\ell(\tilde{x}) \right\}}{\int_{\underline{x}_\ell}^{\phi_\ell^1(x)} \lambda_\ell dF_\ell(\tilde{x}) + \int_{\underline{x}_\ell}^{\phi_\ell^0(x)} (1-\lambda_\ell) dF_\ell(\tilde{x}),} \geq \frac{\lambda_\ell \left\{ \int_{\underline{x}_\ell}^{\psi_\ell^1(x)} dF_\ell(\tilde{x}) \right\}}{\int_{\underline{x}_\ell}^{\psi_\ell^1(x)} dF_\ell(\tilde{x})} = \lambda_\ell,$$

where the equality holds only when $\psi_\ell^1(x) = \psi_\ell^0(x)$, which happens only if the stakeholder is in the mixing region (i.e. $x_\ell < \hat{x}_\ell$). That is, when firm is indifferent between hiring two stakeholders with the same ability but different preferences, then $P_\ell^1(x) = \lambda_\ell$, for relatively small firms $x_{N+1} < x_{N+1}^*(\hat{x}_{\ell,t})$.

For any firm $x_{N+1} > x_{N+1}^*(\hat{x}_{\ell,t})$, we have $\psi_\ell^1(x) > \psi_\ell^0(x)$ and thus $P_\ell^1(x) > \lambda_\ell$. This thus means that, for $x_{N+1} > x_{N+1}^*(\hat{x}_{\ell,t})$. Hence, the share of firms that is matched with a value-driven stakeholders must be larger than λ_ℓ in the region of $[x_{N+1}^*(\hat{x}_{\ell,t}), x_{N+1}]$, $P_\ell^1(x_{N+1}) > \lambda_\ell$ $\forall x_{N+1} \in (x_{N+1}^*(\hat{x}_{\ell,t}), \bar{x}_{N+1})$. Note that, lastly, let x_{N+1}^u denote the highest type that is matched with a value driven stakeholder, and thus $\psi_\ell^1(x_{N+1}^u) = \bar{x}_\ell$. Hence, the measure of top firms that is strictly better off to hire a pecuniary stakeholder is given by

$$\int_{x_{N+1}^u}^{\bar{x}_{N+1}} dF_{N+1}(\tilde{x}) = \Pr\left(z_N \geq (1-c\sigma)\left(\bar{x}_1 \bar{x}_2 \bar{x}_{3...} \bar{x}_N\right)\right)$$

and thus for any $x_{N+1} \geq x_{N+1}^u$, $P_\ell^g(x_{N+1}) = \frac{\lambda_\ell}{(1-F_\ell(x_{N+1}))}$, and $P_\ell^g(x_{N+1})\left(1-F_\ell(x_{N+1})\right) - P_\ell^g(x_{N+1}^u)\left(1-F_\ell(x_{N+1}^u)\right) = 0$. □

A.9 Proof of Proposition 7

Proof. Given that $\hat{x}_{\ell,1} < \hat{x}_{\ell,0}$, for firm x_{N+1} that are matched with stakeholder $x_\ell < \hat{x}_{\ell,1}$, that is, $x_{N+1} < x^*_{N+1}(\hat{x}_{\ell,1})$, then according to Lemma A.8,

$$P^g_{\ell,1}(x) - P^g_{\ell,0}(x) = \lambda_{\ell,1} - \lambda_{\ell,0} = \delta,$$

For firms that were in the mixing region before the shock, but in the separation region after the shock (i.e., $x_{N+1} \in [x^*_{N+1}(\hat{x}_{\ell,1}), x^*_{N+1}(\hat{x}_{\ell,0})]$), then we have

$$P^g_{\ell,1}(x) - P^g_{\ell,0}(x) > \lambda_{\ell,1} - \lambda_{\ell,0} = \delta,$$

where we use the fact that $P^g_{\ell,0}(x) = \lambda_{\ell,0}$ as $x_{N+1} < x^*_{N+1}(\hat{x}_{\ell,0})$. That is, these firms were in the mixing region before the shocks. Lastly, the total changes in the share at the top region $[x^*_{N+1}(\hat{x}_{\ell,0}), \bar{x}_{N+1}]$ is given by

$$\begin{aligned}
&\left\{P^g_{\ell,1}(\bar{x}_{N+1}) - P^g_{\ell,1}(x^*_{N+1}(\hat{x}_{\ell,0}))\right\} - \left\{P^g_{\ell,0}(\bar{x}_{N+1}) - P^g_{\ell,0}(x^*_{N+1}(\hat{x}_{\ell,0}))\right\} \\
=&\delta + P^g_{\ell,0}(x^*_{N+1}(\hat{x}_{\ell,0})) - P^g_{\ell,1}(x^*_{N+1}(\hat{x}_{\ell,0})) \\
<&\delta + \lambda_0 - \lambda_1 < \delta,
\end{aligned}$$

as $P^g_{\ell,0}(x^*_{N+1}(\hat{x}_{\ell,0})) = \lambda_0$ and $P^g_{\ell,1}(x^*_{N+1}(\hat{x}_{\ell,0})) > \lambda_1$, as $x^*_{N+1}(\hat{x}_{\ell,0}) > x^*_{N+1}(\hat{x}_{\ell,1})$. That is, firms on the top has impact lower than δ. □