Sustainable Organizations*

Thomas Geelen† Jakub Hajda‡ Jan Starmans§

May 14, 2023

Abstract

We analyze how stakeholders such as employees, managers, and investors shape organizations when they are pro-social. Our findings challenge the notion that pro-social stakeholders always improve an organization's sustainability. Instead, they demonstrate that conflicts of interest arising from differences in pro-social preferences can result in pro-social stakeholders losing control rights and influence, an effect that ultimately reduces the sustainability of organizations. Our findings shed light on recent trends in stakeholder engagement and provide conditions under which pro-social stakeholders either benefit or harm the sustainability of organizations.

Keywords: Sustainability, ESG, stakeholders, delegation of authority, control rights, corporate governance.

JEL Classifications: D23, G30, L20, Q56.

*We would like to thank Patrick Bolton, Martin Boyer, Alvin Chen, Alex Edmans, Peter Feldhütter, Simon Gervais, Mariassunta Giannetti, Deeksha Gupta, Martin Hellwig, Marcin Kacperczyk, Doron Levit, Nadya Malenko, Ulrike Malmendier, Vincent Maurin, Erwan Morellec, Marcus Opp, Lasse Heje Pedersen, Enrico Perotti, Georgii Riabov, Bogdan Stacescu (discussant), Vladimir Vladimirov, Paul Voss, Paul Whelan, Josef Zechner, and seminar participants at Copenhagen Business School, Finance Theory Group Meeting at the University of Michigan, HEC Montréal, McGill University, the University of Bonn, Johns Hopkins University, the 2022 Young Scholars Nordic Finance Workshop, the University of Amsterdam, and the Stockholm School of Economics for helpful comments. Support from the Danish Finance Institute is gratefully acknowledged.

†Copenhagen Business School and Danish Finance Institute. E-mail: tag.fi@cbs.dk.

‡HEC Montréal. E-mail: jakub.hajda@hec.ca.

§Stockholm School of Economics. E-mail: jan.starmans@hhs.se.

Stakeholders such as employees, managers, and investors are demonstrating a growing commitment to addressing environmental, social, and governance (ESG) issues. In this paper, we investigate how stakeholders' pro-social preferences shape organizations and we challenge the notion that having more pro-social stakeholders always improves an organization's sustainability. The reason is that conflicts of interest can arise with regard to sustainability policies as some stakeholders become more pro-social. Due to these conflicts, stakeholders who are more pro-social can lose their control rights and influence, and this can negatively impact an organization's sustainability. The recent controversy surrounding BlackRock's ESG strategy offers an example of the potential consequences of such conflicts. Several American states have withdrawn investment mandates from BlackRock over the concern that its increasingly important ESG strategy will have a negative impact on investor returns. Consequently, while BlackRock's strategy became more pro-social, its ability to influence the efforts of its portfolio companies to promote sustainability has been reduced.[1]

To investigate how differences in stakeholders' pro-social preferences affect organizations, we incorporate pro-social preferences and a project choice that trades off monetary and social payoffs into a delegation of authority model (e.g., Aghion and Tirole, 1997; Burkart et al., 1997; Stein, 2002). In the model, an organization is composed of two stakeholders that are involved in implementing a project: a principal such as a manager and an agent such as an employee. Since each stakeholder can have pro-social preferences in addition to monetary incentives, the preferred projects of the stakeholders can diverge, leading to a conflict of interest. For example, a car-rental company may have to decide whether to purchase more expensive—but less polluting—electric cars instead of gasoline cars. When selecting the type of car to purchase, stakeholders such as the company's CEO and its fleet manager may disagree about whether it is preferable to favor profitability or environmental concerns.[2]

The stakeholder holding the control rights has the authority to choose the project. However, since one must be informed of the project's payoffs in order to exercise the control rights, the stakeholder may be unable to do so. This means that, even without the control

[1] "Florida to pull $2bn from BlackRock in spreading ESG backlash," Financial Times, 1 December 2022.

[2] While we focus on the trade-off between monetary and social payoffs, our model can be applied to other settings in which an organization's output has multiple dimensions. For example, there may be a trade-off between long-term and short-term profits and heterogeneity in stakeholders' preferences regarding short-term and long-term profits driven by different liquidity needs. However, the model applies particularly well to the trade-off we study because of the non-contractibility of stakeholders' actions, which is especially relevant in the context of organizations' sustainability outcomes.

rights, a stakeholder can exert effective control over the project choice by being better informed than the stakeholder who holds the control rights. For example, a CEO who holds the control rights can decide which type of car to purchase, but choosing a car requires carefully investigating the pros and cons of different models. If the fleet manager understands these pros and cons better than the CEO, then it is best for the CEO to follow the fleet manager's advice regarding the type of car to purchase, and thus to grant the fleet manager effective control over this decision.

The principal faces a trade-off when deciding whether or not to delegate the control rights to the agent. On the one hand, the conflict of interest regarding the project choice makes delegation costly. On the other hand, delegating the control rights can also have benefits, because it can shift the burden of the costly information acquisition to the agent. Consequently, if the principal's and the agent's pro-social preferences differ substantially, then the cost of delegating the control rights outweighs its benefits, and the principal refrains from delegating. Conversely, if the pro-social preferences of the two stakeholders are aligned enough to make the conflict of interest between them less severe, then the principal delegates the control rights to the agent.

The key result of our paper is that an organization's overall sustainability can decrease even if stakeholders become more pro-social because control rights shift from the more pro-social stakeholder to the less pro-social one. There are two cases in which this happens due to the agent becoming more pro-social. The first one occurs when the agent becomes significantly more pro-social than the principal. When this happens, the latter withdraws the control rights because the conflict of interest becomes too severe. The second case occurs when an increase in the agent's pro-social preferences results in a better alignment with the principal's incentives. In this situation, the principal delegates the control rights to the less pro-social agent to save on effort costs. In both cases, the agent becomes more pro-social and the control rights shift from the more pro-social stakeholder to the less pro-social one, harming the organization's sustainability. However, in the first case, the shift in the control rights is due to a more severe conflict of interest, whereas in the second case, it results from a weaker conflict of interest. Yet, in both cases, the organization's sustainability declines. The recent case of BlackRock exemplifies the negative effect that a shift in control rights can have on sustainability outcomes. When some American states withdrew their investment mandates—and thus withdrew their control rights—this reduced BlackRock's capacity to influence the ESG policies of its portfolio companies. By drawing attention to the impact of shifts in control rights, our analysis thus highlights the fact that they can be a significant determinant of the sustainability of organizations.

Interestingly, we show that a shift in the control rights can also benefit an organization's sustainability when it is caused by the principal becoming more pro-social. If the principal is less pro-social than the agent, then this change can cause the principal to delegate the control rights to the more pro-social agent. On the other hand, if the principal is more pro-social than the agent, then this change can result in the principal withdrawing the control rights from the less pro-social agent. In both cases, the control rights shift from the less pro-social to the more pro-social stakeholder and this benefits the organization's sustainability. The decision of hedge fund Engine No. 1 to confront ExxonMobil's management serves as an example of this positive effect. The demands of Engine No. 1 resulted in the election of three new directors to ExxonMobil's board of directors. This resulted in a shift in the control rights and posed a challenge to the company's existing strategy because these new directors pushed for a transition to renewable energy.[3]

We also demonstrate that changes in pro-social preferences play an important role when the control rights remain unchanged. On the one hand, these changes alter stakeholders' preferred projects. On the other hand, these changes impact stakeholders' incentives to become informed about the projects' payoffs, which can strengthen or weaken the influence of the more pro-social stakeholders. For example, we show that making the agent more pro-social relative to the principal benefits the organization's sustainability when control rights remain unchanged. This happens both because the agent's preferred project becomes more pro-social and because the agent has a greater incentive to acquire information, which results in a shift in effective control from the less pro-social principal to the more pro-social agent.

In sum, we provide a positive theory of stakeholder governance that sheds light on how pro-social stakeholders affect organizations. Our framework makes it possible to identify situations where strengthening stakeholders' pro-social preferences harms an organization's sustainability. In particular, our results show that although pro-social principals benefit an

[3] "Signs of change at ExxonMobil a year after hedge fund proxy fight," Financial Times, 24 May 2022.

organization's sustainability, it is not always true that pro-social agents do. In other words, stronger pro-social preferences enhance organizational sustainability when implemented from the top-down, but may yield unintended consequences when operating from the bottom-up. Moreover, we demonstrate that even minor changes in pro-social preferences can have a significant effect on an organization's sustainability because they can alter the allocation of the control rights.

We extend our model to examine the role of ESG-linked compensation and agent selection (e.g., hiring, internal promotions, investor selection) in shaping an organization's sustainability. We show that while social compensation can mitigate the conflict of interest among stakeholders by incentivizing them to choose projects with a higher social payoff, it may not necessarily contribute to the overall sustainability of the organization. This negative effect arises when an increase in the agent's social compensation leads to a shift in the control rights, which can harm the organization's sustainability. Finally, we show that when choosing among agents with varying levels of pro-social preferences, the principal is inclined to select an agent who is more but not overly pro-social. The rationale is that a more pro-social agent has greater incentives to exert effort, which is beneficial for the principal. However, if the agent were too pro-social, the conflict of interest would become too severe, which would negatively impact the principal. We also show that the selection of a more but not overly pro-social agent benefits an organization's sustainability.

Our model has empirical implications for different types of relationships between stakeholder including managers and employees, entrepreneurs and investors, investors and managers, as well as company boards of directors and CEOs. For example, our model implies that making stakeholders more pro-social can benefit or harm an organization's sustainability, which is in line with the mixed empirical evidence regarding the impact of investors on the sustainability footprint of firms and other organizations (e.g., Kim et al., 2022; Heath et al., 2021; Di Giuli and Kostovetsky, 2014; Chen et al., 2020; Huang et al., 2021). Our results also provide a theoretical underpinning for the increasing prevalence of stakeholder activism and engagement—stakeholders trying to obtain effective control—in relation to sustainability issues (e.g., Eccles and Klimenko, 2019). Section IV contains a detailed discussion of the model's implications for CEO authority and retention, board composition and dynamics, shareholder proposals, and shareholder engagement and activism on ESG issues.

Our paper provides a theory of stakeholder society (e.g., Tirole, 2001; Allen et al., 2015; Magill et al., 2015) when stakeholders have pro-social preferences and contributes to several strands of the literature. First, our analysis adds to the literature in organizational economics that studies control rights and stakeholders' incentives to acquire information to exercise these rights (e.g., Aghion and Tirole, 1997; Burkart et al., 1997; Stein, 2002).[4] We add to the existing literature by considering organizations that generate both monetary and social payoffs, and by endowing stakeholders with pro-social preferences. Introducing these elements allows us to uncover how stakeholders influence organizations' sustainability and to provide testable implications that are absent in prior literature.[5] For example, our analysis reveals that more pro-social stakeholders may negatively impact the sustainability of organizations due to a shift in control rights to less pro-social stakeholders. Importantly, this outcome occurs when agents become more pro-social but not when principals become more pro-social. Additionally, our analysis yields novel implications for the interaction between stakeholders' pro-social preferences, compensation, and hiring.

Second, our paper contributes to the theoretical literature on corporate governance by studying the impact of pro-social stakeholders. See Malenko (2022) for a survey on the literature on corporate governance. This is particularly important in light of increasing empirical evidence highlighting the importance of sustainability concerns in the context of corporate governance (e.g, Di Giuli and Kostovetsky, 2014; Dimson et al., 2015; McCahery et al., 2016; Hoepner et al., 2018; Dyck et al., 2019; Krueger et al., 2020; Dasgupta et al., 2021). In contrast to the small but growing theoretical literature on the impact of pro-social stakeholders on corporate governance (Matsusaka and Shu, 2021; Gollier and Pouget, 2022; Levit et al., 2022), our paper examines how control rights and the allocation of effective control influence the sustainability of organizations. We demonstrate that control rights and

[4]A related literature studies how control rights affect the communication between agents in organizations (e.g., Dessein, 2002; Harris and Raviv, 2005; Alonso et al., 2008; Grenadier et al., 2016). See Bolton and Dewatripont (2013) for a survey of the literature on authority in organizations. Delegation-of-authority models have been used to study questions in, amongst others, corporate finance, corporate governance, and corporate culture (e.g., Burkart et al., 1997; Van den Steen, 2010a,b; Chen, 2022).

[5]For example, our framework allows us to address the question of whether control rights are allocated to more or less pro-social stakeholders, which is crucial for understanding the effect on the sustainability of organizations. This is a question that standard delegation-of-authority models, such as Aghion and Tirole (1997), cannot address. In addition, our analysis includes rich interactions between principals and agents that are not present in their work. Specifically, changes in the agent's pro-social preferences not only impact the agent's project choice and utility, but also effect the principal's utility when the agent has effective control.

effective control are crucial dimensions for understanding the effects of pro-social stakeholders on organizations.

Finally, we contribute to the growing theoretical literature on the impact of pro-social stakeholders on organizations. Previous research has primarily focused on investors and firms (e.g., Heinkel et al., 2001; Chowdhry et al., 2019; Morgan and Tumlinson, 2019; Landier and Lovo, 2020; Green and Roth, 2021; Roth, 2021; Broccardo et al., 2022; Gupta et al., 2022; Hart and Zingales, 2022; Oehmke and Opp, 2022).[6] In contrast, we develop a theory of stakeholder control and engagement that applies to a broader range of stakeholders and organizations. As stakeholders increasingly demand that organizations address ESG issues, it is important to understand how the interactions between stakeholders influence the sustainability of organizations.

I Model

We consider an organization composed of two risk-neutral stakeholders: a principal P and an agent A. The principal is the controlling stakeholder in the organization. For example, the principal is a manager of a firm and the agent is an employee. There exists a set of projects that differ in terms of their social and monetary payoffs and the organization can implement one of the projects. There are three dates without time discounting. At time zero, the principal decides whether to delegate the control rights. At time one, both the principal and the agent decide how much effort to exert to become informed about the projects' payoffs. At time two, a project may be implemented if at least one stakeholder is informed. We describe the model in more detail below.

The output of the organization is a pair (π, s), where π is the monetary payoff and s is the social payoff. We refer to a payoff pair (π, s) as a project. The organization has one unit of initial resources which can be employed to produce the monetary and social payoffs. Let $\iota \in [0, 1]$ denote the investment in the social payoff, then $1 - \iota$ is the investment in the monetary payoff. There exists a production technology which generates a monetary payoff of $\pi = \sqrt{1 - \iota}$ and a social payoff of $s = \sqrt{\iota}$. Consequently, the relevant set of projects is

[6] See Gillan et al. (2021) for a survey. A related literature studies the asset pricing implications of socially responsible investors (e.g., Pástor et al., 2021; Pedersen et al., 2021; Pedersen and Feldhütter, 2022).

given by

$$\mathcal{P} = \left\{ \left(\sqrt{1-\iota}, \sqrt{\iota}\right) \middle| \iota \in [0,1] \right\}.$$

There exists no savings technology and thus the organization generates a zero monetary payoff and a zero social payoff if the initial resources are not employed.

The two stakeholders face an informational friction because without acquiring additional information there is a chance of generating highly negative payoffs. As a result, if neither stakeholder is informed, then no project is implemented.[7] Both stakeholders can exert costly effort to become informed about the production technology. Specifically, a stakeholder $j \in \{P, A\}$ chooses a probability $q_j \in [0,1]$, which corresponds to effort exerted on learning about the production technology. The stakeholder j's private cost of effort is $\frac{\phi_j}{2} q_j^2$. If informed, the stakeholder j can choose a project (π, s) from the set $\mathcal{P}$. The stakeholders make their effort choices simultaneously at time one and the outcomes of the principal's and the agent's effort choices at time two are independent. An alternative interpretation of the informational friction is that there exists a search cost to identify relevant projects.

It is well documented that some stakeholders have pro-social preferences and that they may differ in terms of these preferences (e.g., Hong and Kacperczyk, 2009; List, 2009; Hong and Kostovetsky, 2012; Gibson et al., 2021). Thus, the stakeholder j's utility from implementing a project (π, s) is

$$u_j(\pi, s) = \beta_j \pi + \gamma_j s,$$

where $\gamma_j \geq 0$ captures the stakeholder's pro-social preferences and $\beta_j > 0$ represents the monetary incentives. To ensure an interior equilibrium in the stakeholders' effort choices, we assume that effort is sufficiently costly: $\phi_j > \max_{(\pi,s)\in\mathcal{P}} u_j(\pi, s)$, $j \in \{P, A\}$.

Finally, at time zero, the principal decides on the delegation of the control rights $d \in \{P, A\}$ over the organization's project choice. The principal either retains the control rights, $d = P$, or delegates them to the agent, $d = A$. The stakeholder holding the control rights has the authority to choose the project at time two but can also delegate the organization's project choice to the other stakeholder ex post. We refer to the stakeholder whose preferred

[7] What matters in our framework is that the principal is better off with the agent's preferred project than with a random project and vice versa for the agent. The existence of projects with highly negative payoffs is a sufficient condition for this assumption to hold.

project is implemented as the stakeholder holding effective control.[8,9]

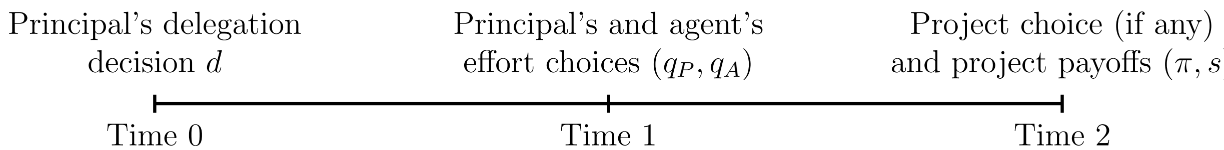


Figure 1: **Model timeline.**

Figure 1 presents the timeline of the model. First, the principal decides whether to delegate the control rights d. Second, the two stakeholders decide how much effort to exert (q_P, q_A). Finally, the project is chosen and implemented (if any) and payoffs realize.

II Equilibrium Analysis

We solve the model by backward induction. We first determine each stakeholder's preferred project. Next, we determine the principal's and agent's effort choices. Finally, we characterize the principal's delegation decision.

A Project Choice

At date two, a stakeholder j chooses the preferred project by maximizing utility, that is,

$$\max_{\iota \in [0,1]} u_j \left(\sqrt{1-\iota}, \sqrt{\iota} \right),$$

which yields $\iota_j = R_j$, where $R_j = \frac{\gamma_j^2}{\gamma_j^2 + \beta_j^2} \in [0, 1)$ is what we refer to as the stakeholder's *relative pro-social preferences*. Thus, the stakeholder j's preferred project is given by $(\pi_j, s_j) = \left(\sqrt{1-R_j}, \sqrt{R_j} \right)$.

Monetary incentives and pro-social preferences affect a stakeholder j's project choice by means of the relative pro-social preferences R_j. A stakeholder with stronger relative pro-

[8]If the principal retains the control rights at time zero, then this is equivalent to postponing the delegation of the control rights until after the information acquisition stage.

[9]We assume that the principal cannot renege on the delegation decision at time two. This assumption clearly holds if the control rights are contractually agreed on, for example, by giving outside investors voting equity. If this is not the case, the assumption can be justified if the principal incurs a reputational cost larger than $u_P(\pi_P, s_P) - u_P(\pi_A, s_A)$ when reneging on the promise to delegate the control rights. Baker et al. (1999) micro-found this reputational cost in a repeated delegation-of-authority model.

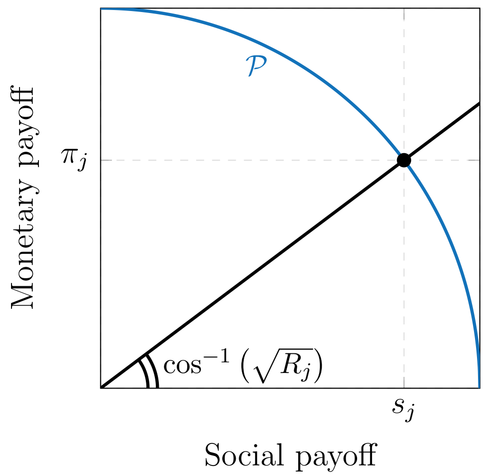


Figure 2: **Set of projects and a stakeholder's preferred project.** The figure plots the set of projects $\mathcal{P}$ and a stakeholder j's preferred project (π_j, s_j) for a given level of the stakeholder's relative pro-social preferences R_j.

social preferences is more willing to accept a lower monetary payoff to generate a higher social payoff and therefore invests more of the organization's initial resources in the social payoff and less in the monetary payoff. For example, an employee with stronger pro-social preferences is more willing to accept a lower wage or bonus if the organization generates a higher social payoff. This implication is consistent with the findings of Krueger et al. (2022) who show that workers in more sustainable sectors earn lower wages. The authors attribute this wage gap to workers' preferences for environmental sustainability. Similarly, an investor with stronger pro-social preferences is willing to accept a lower financial return in exchange for a higher social return, consistent with evidence in, for example, Riedl and Smeets (2017), Bonnefon et al. (2019), and Heeb et al. (2022).

When the two stakeholders j and j' have the same relative pro-social preferences, $R_j = R_{j'}$, then their preferred projects are the same: $(\pi_j, s_j) = (\pi_{j'}, s_{j'})$. In particular, without pro-social preferences, $\gamma_j = \gamma_{j'} = 0$, their preferred projects coincide. Increasing the wedge between the stakeholders' relative pro-social preferences, $|R_j - R_{j'}|$, makes their preferred projects differ more and therefore makes the conflict of interest regarding project choice more severe. As we show below, this conflict of interest is crucial in determining who holds the control rights and therefore in shaping the organization's sustainability.

B Effort

Given the principal's and agent's preferred projects, we can determine the stakeholders' expected utilities at time one, which in turn determine their effort choices. Figure 3 summarizes the project choice at time two depending on the allocation of the control rights and the information of the principal and the agent. For example, assume that the principal holds the control rights, which is depicted Figure 3a. In that case, the principal implements the preferred project (π_P, s_P) when informed. When the principal is uninformed while the agent is informed, then the principal follows the agent's recommendation and implements the agent's preferred project (π_A, s_A). This means that the agent has effective control.[10] If neither stakeholder is informed, then no project is implemented due to the risk of generating highly negative payoffs. The other case in which the agent holds the control rights is similar, the only difference being that the agent's preferred project is implemented if both stakeholders are informed.

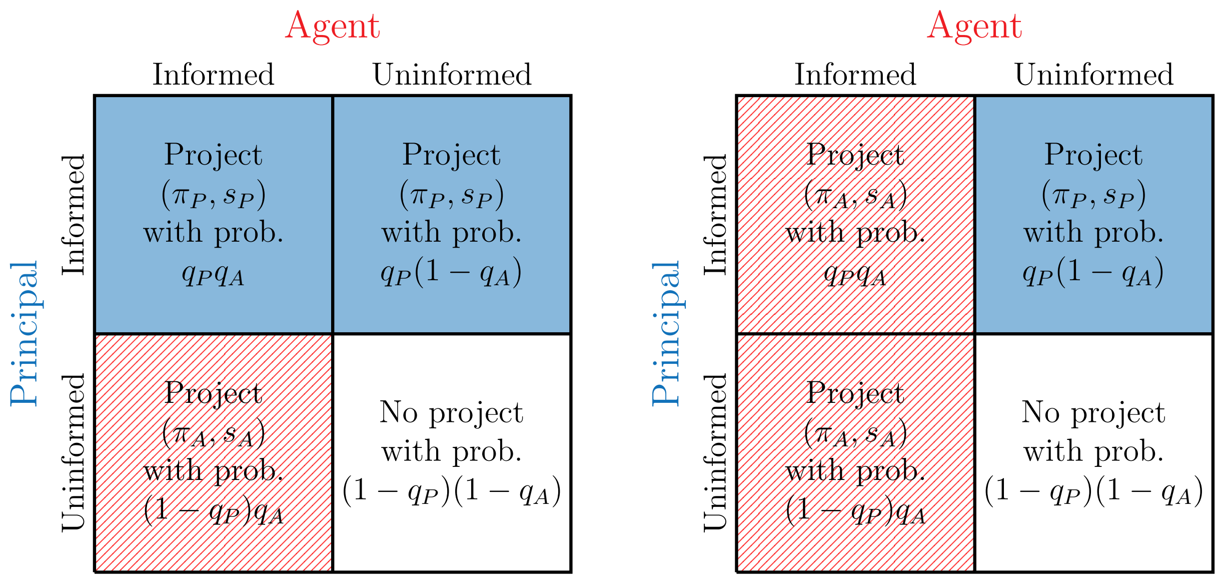
The stakeholder j's expected utility at time one is given by

$$U_j(q_P, q_A, d) = \begin{cases} q_P u_j(\pi_P, s_P) + (1 - q_P) q_A u_j(\pi_A, s_A) - \frac{\phi_j}{2} q_j^2, & \text{if } d = P, \\ (1 - q_A) q_P u_j(\pi_P, s_P) + q_A u_j(\pi_A, s_A) - \frac{\phi_j}{2} q_j^2, & \text{if } d = A. \end{cases}$$

The principal has effective control with probability q_P when holding the control rights, $d = P$, while the agent has effective control with probability $(1 - q_P)q_A$, which occurs when the agent is informed, but the principal is not. On the other hand, if the principal delegates the control rights to the agent, $d = A$, the probability of having effective control for the principal decreases to $(1 - q_A)q_P$, while the probability for the agent increases to q_A.

The stakeholders choose their effort levels simultaneously. As we show in Lemma 3 in Appendix C, the equilibrium effort choices $(q_P(d), q_A(d))$ conditional on the delegation decision d, are determined by the two first-order conditions of the principal's and agent's expected utilities and satisfy $(q_P(d), q_A(d)) \in (0, 1)^2$. This implies that there is a positive probability that each stakeholder determines the organization's project choice and thus has effective control. As a result, both the principal's and the agent's relative pro-social preferences affect

[10] The principal follows the agent's recommendation because $u_P(\pi_A, s_A) > 0$. The stakeholders' relative pro-social preferences satisfy $R_j < 1$ because $\beta_j > 0$. As a result, we have $\pi_j > 0$ while $s_j \geq 0$ and therefore $u_j(\pi_{j'}, s_{j'}) \geq \beta_j \pi_{j'} > 0$, where j and j' denote the two stakeholders.



(a) Principal holds control rights: $d = P$. (b) Agent holds control rights: $d = A$.

Figure 3: **Control rights and effective control.** This figure summarizes which project is undertaken (if any) at time two and the probability of the different cases as a function of the allocation of the control rights and the principal's and agent's effort. The background color indicates who holds effective control, where blue-filled (red-shaded) indicates that the principal (agent) has effective control.

the organization's expected social payoff.

The best response function of the stakeholder j is informative about the stakeholder's effort choice and is given by

$$B_j(q_{j'}, d) = \frac{(1-q_{j'})u_j(\pi_j, s_j) + \mathbb{I}_{\{d=j\}} q_{j'} \Delta u_j}{\phi_j},$$

where j' denotes the other stakeholder and $\Delta u_j = u_j(\pi_j, s_j) - u_j(\pi_{j'}, s_{j'}) \geq 0$. The best response function reveals three important properties of the stakeholders' effort choices. First, the stakeholders' effort choices are strategic substitutes because $\frac{\partial B_j(q_{j'},d)}{\partial q_{j'}} < 0$. Second, a stakeholder's monetary incentives and pro-social preferences affect the incentives to exert effort through the utility $u_j(\pi, s)$ received from implementing the preferred project (π_j, s_j) and the other stakeholder's preferred project $(\pi_{j'}, s_{j'})$. Finally, holding the control rights increases the incentives to exert effort. This is because the stakeholder with the control rights can implement the preferred project in the case when both stakeholders are informed.

We can now translate the effort choices into the allocation of effective control. Specifically, we define the principal's allocation of effective control as the probability that the principal determines the organization's project choice, conditional on a project being implemented, denoted by $e_P(d)$.[11] The agent's allocation of effective control is $e_A(d) = 1 - e_P(d)$.

Throughout the paper, we mostly focus on the cases in which one of the stakeholders has no pro-social preferences and consider an increase in the other stakeholder's pro-social preferences. As Lemma 1 shows, in the case where the principal has no pro-social preferences, $\gamma_P = 0$, an increase in the agent's pro-social preferences γ_A unambiguously increases the agent's effort and effective control.

Lemma 1 (Pro-social Preferences, Effort, and Effective Control)**.** *When the principal has no pro-social preferences, $\gamma_P = 0$, then an increase in the agent's pro-social preferences γ_A leads to an increase in the agent's effort, $\frac{\partial q_A(d)}{\partial \gamma_A} \geq 0$, and to a decrease in the principal's effort, $\frac{\partial q_P(d)}{\partial \gamma_A} \leq 0$. This substitution in effort between the stakeholders translates into a shift in effective control from the principal to the agent, that is, $\frac{\partial e_A(d)}{\partial \gamma_A} \geq 0$ and $\frac{\partial e_P(d)}{\partial \gamma_A} \leq 0$.*

An increase in the agent's pro-social preferences affects the agent's effort choice in two ways. First, becoming more pro-social increases the agent's level of utility and incentives to exert effort.[12] Second, higher pro-social preferences of the agent render the conflict of interest with the principal more severe, which can further increase the agent's incentives to exert effort. Even though a more severe conflict of interest can also increase the principal's incentive to exert effort, the direct effect on the agent's utility always dominates the indirect effect on the principal's utility.

Due to the symmetry of the model, we obtain the same result when the agent has no pro-social preferences, $\gamma_A = 0$, in that an increase in the principal's pro-social preferences

[11] We focus on the probability conditional on a project being implemented and not on the unconditional probability because in practice projects that aren't undertaken often cannot be observed in the data. Both stakeholders are uninformed with probability $(1-q_P(d))(1-q_A(d))$. As a result, the probability of a project being implemented is given by $q_P(d) - q_P(d)q_A(d) + q_A(d)$ and we get $e_P(P) = \frac{q_P(P)}{q_P(P)-q_P(P)q_A(P)+q_A(P)}$ and $e_P(A) = \frac{(1-q_A(A))q_P(A)}{q_P(A)-q_P(A)q_A(A)+q_A(A)}$.

[12] One of the key forces through which an increase in a stakeholder's pro-social preferences affects the effort level is by increasing the stakeholder's utility. Intuitively, strengthening the pro-social preferences of a stakeholder means that the stakeholder cares more about the social payoff. In particular, we argue that absolute shifts in pro-social preferences rather than relative shifts from monetary incentives to pro-social preferences are the right way to examine changes in stakeholders' pro-social preferences. For example, if an employee receives a salary and becomes more environmentally conscious, it does not mean that the employee cares less about monetary income, only that the environmental impact of the company becomes more important.

γ_P leads to a substitution in effort and effective control from the agent to the principal. In Appendix A, we study in more detail how changes in stakeholders' monetary incentives and pro-social preferences impact the equilibrium effort levels and therefore effective control when both stakeholders have pro-social preferences.[13]

Given that changes in the stakeholders' pro-social preferences alter the project choices and the allocation of effective control, it is crucial to understand how those changes jointly affect the organization's sustainability. To this end, we examine the expected social payoff conditional on the delegation decision and on a project being implemented,

$$\mathbb{E}_0\left[\tilde{s}|\tilde{\pi} > 0, d\right] = e_P(d)s_P + e_A(d)s_A,$$

where $\tilde{\pi}$ and $\tilde{s}$ are the random monetary and social payoffs and where $\mathbb{E}_0$ denotes the expectation at time zero. We refer to the organization's expected social payoff conditional on a project being implemented, $\mathbb{E}_0\left[\tilde{s}|\tilde{\pi} > 0\right]$, as the *organization's sustainability.* Intuitively, the higher the expected social payoff is, the higher the organization would score on sustainability KPIs and the more sustainable it would be deemed. Note that this result does not imply that more sustainable organizations are always desirable from a welfare perspective because higher social payoffs come at the cost of lower monetary payoffs.

The effect of changing the agent's pro-social preferences on the organization's sustainability, conditional on the delegation decision, is

$$\frac{\partial \mathbb{E}_0\left[\tilde{s}|\tilde{\pi} > 0, d\right]}{\partial \gamma_A} = \underbrace{\frac{\partial e_A(d)}{\partial \gamma_A}(s_A - s_P)}_{\Delta \text{ Effective Control}} + \underbrace{e_A(d)\frac{\partial s_A}{\partial \gamma_A}}_{\Delta \text{ Project Choice}}.$$

There are two effects at play. First, making the agent more pro-social changes effective control in the organization. This effect has a positive influence on the organization's sustainability when effective control shifts to the most pro-social stakeholder. The second effect is that making the agent more pro-social tilts the agent's preferred project towards one with a higher social payoff, which has a positive influence on the organization's sustainability.

Proposition 1 (Pro-social Preferences and Organization's Sustainability)**.** *When the prin-*

[13] The key difference as compared to the case considered here is that an increase in the agent's pro-social preferences can also make the conflict of interest with the principal less severe, which can decrease the agent's incentive to exert effort.

cipal has no pro-social preferences, $\gamma_P = 0$, then an increase in the agent's pro-social preferences γ_A leads to a shift in effective control to the agent, $\frac{\partial e_A(d)}{\partial \gamma_A} \geq 0$, who chooses a more pro-social project, $\frac{\partial s_A}{\partial \gamma_A} \geq 0$. As a result, the organization's sustainability, conditional on the delegation decision, is increasing in the agent's pro-social preferences:

$$\frac{\partial \mathbb{E}_0\left[\tilde{s} | \tilde{\pi} > 0, d\right]}{\partial \gamma_A} \geq 0.$$

When taking the delegation decision as given, Proposition 1 demonstrates that as the agent becomes more pro-social, then the organization becomes more sustainable. In this case, both the change in effective control and the change in project choice increase the organization's sustainability. Note that the same result holds when the agent has no pro-social preferences, $\gamma_A = 0$, and the principal's pro-social preferences γ_P become stronger. That is, the organization's sustainability increases.

While Proposition 1 shows that more pro-social stakeholders can make an organization more sustainable, this result takes the delegation decision as given. However, this decision is taken by the principal and it can change as stakeholders become more pro-social. We study this last crucial step in the following section.

C Delegation of Control Rights

The principal decides whether to delegate the control rights, taking into account the future actions of both stakeholders, that is,

$$\max_{d \in \{P, A\}} U_P(q_P(d), q_A(d), d),$$

where $q_P(d)$ and $q_A(d)$ are the stakeholders' effort choices.

We first study its impact on the effort choices of the stakeholders.

Lemma 2 (Control Rights and Effort)**.** *Allocating the control rights to a stakeholder increases the stakeholder's effort and reduces the other stakeholder's effort, that is, $q_P(P) \geq q_P(A)$ and $q_A(A) \geq q_A(P)$.*

As Lemma 2 shows, allocating the control rights to a stakeholder increases the stakeholder's effort. This substitution in effort between the stakeholders translates into a shift in effective control, that is, $e_P(P) \geq e_P(A)$ and $e_A(A) \geq e_A(P)$. The reason for the increase is that holding the control rights increases the likelihood of having effective control, which results in a stronger incentive to exert effort.

To study the principal's delegation decision, we define the wedge in the principal's utility gained from retaining rather than delegating the control rights as

$$\Delta U_P = U_P(q_P(P), q_A(P), P) - U_P(q_P(A), q_A(A), A).$$

In particular, the principal delegates the control rights, $d = A$, if $\Delta U_P < 0$ and retains the control rights, $d = P$, if $\Delta U_P > 0$.

Proposition 2 (Irrelevance of Control Rights)**.** *When the stakeholders have the same relative pro-social preferences, $R_P = R_A$, then the delegation decision d does not affect their effort choices, $(q_P(P), q_A(P)) = (q_P(A), q_A(A))$, and expected payoffs. As a result, $\Delta U_P = 0$.*

Proposition 2 shows that the principal's expected utility does not depend on the delegation decision if there is no conflict of interest between the stakeholders. The result highlights that both the pro-social preferences and the wedge in stakeholders' relative pro-social preferences are necessary to make the delegation decision relevant.

To understand how pro-social preferences affect the delegation decision, note that the wedge ΔU_P can be rewritten as

$$\begin{aligned} \Delta U_P = & \underbrace{\mathbb{P}_0(\tilde{\pi} > 0 | d = P)\mathbb{E}_0\left[u_P(\tilde{\pi}, \tilde{s}) | \tilde{\pi} > 0, d = P\right] - \frac{\phi_P}{2} q_P^2(P)}_{\text{Expected utility when } d = P} \\ & - \underbrace{\left(\mathbb{P}_0(\tilde{\pi} > 0 | d = A)\mathbb{E}_0\left[u_P(\tilde{\pi}, \tilde{s}) | \tilde{\pi} > 0, d = A\right] - \frac{\phi_P}{2} q_P^2(A)\right)}_{\text{Expected utility when } d = A}. \end{aligned} \tag{1}$$

In Equation (1), we express the principal's expected utility as the probability that a project is implemented times the expected utility conditional on a project being implemented, minus the effort cost. The equation shows that the delegation decision affects the expected utility in three ways. The first effect of delegating the control rights to the agent is that it impacts the probability of a project being undertaken, which we refer to as the *project implementation* *effect*. Thus, if

$$\mathbb{P}_0(\tilde{\pi} > 0|d = A) > \mathbb{P}_0(\tilde{\pi} > 0|d = P),$$

then the probability of a project being implemented is higher when the agent holds the control rights. This, in turn, is beneficial to the principal. Intuitively, while the principal's expected utility is lower when the agent holds the control rights, delegating the control rights to the agent may increase the probability that a project gets implemented. In other words, delegating the control rights to the agent may reduce the risk of no project being implemented.

The second effect of delegating the control rights to the agent is an *effort cost effect*. The principal exerts less effort and the agent exerts more effort, which lowers the principal's effort cost by

$$\frac{\phi_P}{2}\left(q_P^2(P) - q_P^2(A)\right) \geq 0,$$

and always gives the principal an incentive to delegate control rights.

The third effect of delegating the control rights to the agent is that it alters the likelihood that each stakeholders' preferred project is undertaken. A change in the control rights increases the agent's effort and reduces the principal's effort, which strengthens the agent's effective control. As a consequence, we have

$$\mathbb{E}_0\left[u_P(\tilde{\pi}, \tilde{s})|\tilde{\pi} > 0, d = P\right] \geq \mathbb{E}_0\left[u_P(\tilde{\pi}, \tilde{s})|\tilde{\pi} > 0, d = A\right].$$

Intuitively, delegating the control rights to the agent means that the agent's preferred project is relatively more likely to be implemented, which reduces the principal's expected utility. This effect, which we refer to as the *project selection effect*, discourages the principal from delegating the control rights to the agent.

Proposition 3 (Relative Pro-social Preferences and Control Rights)**.** *Taking as given the principal's monetary incentives β_P and pro-social preferences γ_P, if the wedge in relative pro-social preferences, $|R_P - R_A|$, is positive but sufficiently small, then the principal delegates the control rights to the agent: $d = A$. If the wedge is relatively large, then the principal retains the control rights: $d = P$.*

Intuitively, if the conflict of interest between the principal and the agent is minor, then the project selection effect is small. That is, the shift in effective control to the agent resulting from the delegation of control rights only leads to a small loss in expected utility for the principal. In this case, the value the principal obtains from the agent's increased effort and from the lower effort cost dominates the project selection effect. Therefore, the principal delegates the control rights to the agent. In contrast, if the conflict of interest is severe, then the project selection effect becomes larger and the principal retains the control rights.

We can fully characterize the delegation decision when one of the stakeholders has no pro-social preferences. That is, when $\gamma_P = 0$ and ϕ_A is sufficiently large, we obtain the result in Proposition 4.[14]

Proposition 4 (Principal without Pro-social Preferences and Control Rights)**.** *When the principal has no pro-social preferences, $\gamma_P = 0$, and $\phi_A > \hat{\phi}_A$, where $\hat{\phi}_A$ is defined in the appendix, then there exists a threshold $\hat{\gamma}_A > 0$ such that the principal delegates the control rights when $\gamma_A \in (0, \hat{\gamma}_A)$ and retains the control rights when $\gamma_A > \hat{\gamma}_A$.*

Recall from Proposition 1 that the organization's sustainability, conditional on the delegation decision, is increasing in γ_A when $\gamma_P = 0$. The crucial insight from Proposition 4 is that increasing the agent's pro-social preferences can lead to a withdrawal of the control rights from the agent. The following result shows that this has negative consequences for the organization's sustainability.

Proposition 5 (Principal without Pro-social Preferences and Organization's Sustainability)**.** *When the principal has no pro-social preferences, $\gamma_P = 0$, and $\phi_A > \hat{\phi}_A$, then an increase in the agent's pro-social preferences γ_A increases the organization's sustainability for all $\gamma_A \neq \hat{\gamma}_A$, where $\hat{\phi}_A$ and $\hat{\gamma}_A$ are the thresholds from Proposition 4, that is,*

$$\forall \gamma_A \neq \hat{\gamma}_A, \quad \frac{\partial \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]}{\partial \gamma_A} \geq 0.$$

At $\hat{\gamma}_A$, the organization's sustainability decreases discontinuously, that is,

$$\lim_{\gamma_A \uparrow \hat{\gamma}_A} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0] > \lim_{\gamma_A \downarrow \hat{\gamma}_A} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0].$$

[14] The condition $\phi_A > \hat{\phi}_A$ in Proposition 4 ensures that if γ_A gets sufficiently large, then the principal wants to retain the control rights.

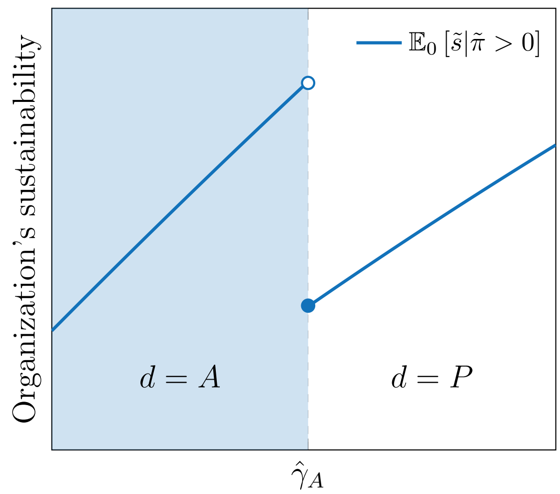


Figure 4: **Agent's pro-social preferences and organization's sustainability.** The figure plots the organization's sustainability, $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$, as a function of the agent's pro-social preferences γ_A. If $\gamma_A < \hat{\gamma}_A$, the principal delegates the control rights and the principal retains the control rights if $\gamma_A > \hat{\gamma}_A$.

Figure 4 shows the total effect of increasing the agent's pro-social preferences from Proposition 5, taking into account the endogenous delegation decision by the principal. The plot highlights the downward jump in the organization's sustainability as the principal withdraws the control rights.[15] For example, if the CEO becomes more socially responsible, they may lose their control rights and therefore have less influence on the firm, which can ultimately harm the firm's sustainability.[16]

In many real-life applications, control rights are discrete. For example, the control of a firm can change around the majority-voting threshold. However, our result in Proposition 5, which demonstrates that an increase in the agent's pro-social preferences can reduce the organization's sustainability due to a withdrawal of the control rights, does not rely on the discrete nature of changing the control rights. In Appendix B we extend our framework to a continuum of projects and control rights. By doing so, we demonstrate that it is not crucial

[15]Note that the result is not driven by the fact that we consider the organization's sustainability as the conditional expectation $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$. For the unconditional expectation $\mathbb{E}_0[\tilde{s}]$, we get that $\mathbb{E}_0[\tilde{s}|d = A] = q_A(A)s_A$ and $\mathbb{E}_0[\tilde{s}|d = P] = q_A(P)(1 - q_P(P))s_A$. Lemmas 1 and 2 imply that each of these expectations is non-decreasing in γ_A and at $\hat{\gamma}_A$, where the control rights shift, we have that $\mathbb{E}_0[\tilde{s}|d = A] > \mathbb{E}_0[\tilde{s}|d = P]$.

[16]Given the symmetry in our model, we can interchange the monetary and social preferences and the payoffs in all formal results to study changes in stakeholders' monetary incentives. Note that in the model we assume that $\beta_P > 0$ and $\beta_A > 0$ so for the interchanged results we would assume that $\gamma_P > 0$ and $\gamma_A > 0$. For example, for the result in Proposition 5 this would imply that as the agent's monetary incentives increase, the organization's expected profitability conditional on a project being undertaken jumps down at $\hat{\beta}_A$.

that the control rights are discrete. Rather, it is important that strengthening the agent's pro-social preferences leads to a significant reduction in the agent's control rights, which negatively impacts the organization's sustainability.

As we show in Proposition 6, even when the principal has pro-social preferences, a change in the control rights that results from an increase in the agent's pro-social preferences harms the organization's sustainability.

Proposition 6 (Agent's Pro-Social Preferences, Changes in Control Rights, and Organization's Sustainability)**.** *For a given set of preference parameters* $\{\beta_P, \gamma_P, \beta_A\}$*, assume that the delegation set is convex:* $\{\gamma_A | d = A\} = [\underline{\gamma}_A, \overline{\gamma}_A]$*. Further assume that the* γ_A *that implies no conflict of interest,* $\{\gamma_A | R_P = R_A\}$*, is part of the feasible parameter space.*[17] *If the allocation of the control rights changes at* $\gamma'_A \in \{\underline{\gamma}_A, \overline{\gamma}_A\}$*, then the organization's sustainability decreases discontinuously:*

$$\lim_{\gamma_A \uparrow \gamma'_A} \mathbb{E}_0[\tilde{s} | \tilde{\pi} > 0] > \lim_{\gamma_A \downarrow \gamma'_A} \mathbb{E}_0[\tilde{s} | \tilde{\pi} > 0].$$

The control rights can change at the boundary of the delegation set $\{\underline{\gamma}_A, \overline{\gamma}_A\}$. If they change at $\underline{\gamma}_A$, then the agent is less pro-social than the principal.[18] At this threshold, the principal starts delegating the control rights to the agent because the conflict of interest becomes weaker when the agent's pro-social preferences increase. At $\overline{\gamma}_A$, the agent is more pro-social than the principal. At this threshold, the principal withdraws the control rights from the agent because the conflict of interest becomes more severe.

In both cases, the control rights shift from the more pro-social stakeholder to the less pro-social one, harming the organization's sustainability. This highlights our main result that more pro-social stakeholders can harm an organization's sustainability. Importantly, in the first case, the shift in the control rights is due to a less severe conflict of interest, whereas in the second case it results from a more severe conflict of interest. Yet, in both cases, the organization's sustainability declines.

We now turn to analyzing the case in which the principal becomes more pro-social. As

[17] In Proposition 6, we assume that if the principal is indifferent between delegating or retaining the control rights, $\Delta U_P = 0$, then the principal delegates the control rights.

[18] The control rights only change at $\underline{\gamma}_A$ if it is positive, otherwise the principal always delegates the control rights if γ_A is close to or equals zero. Furthermore, we can show that if $\underline{\gamma}_A > 0$ then $\underline{\gamma}_A < \{\gamma_A | R_P = R_A\}$ and therefore the agent is less pro-social than the principal at $\underline{\gamma}_A$.

we show below, changes in the control rights lead to an upward jump in the organization's sustainability, in sharp contrast to the case in which the agent becomes more pro-social. In Proposition 7, we first study the principal's delegation decision when the agent has no pro-social preferences.

Proposition 7 (Agent without Pro-social Preferences and Control Rights)**.** *When the agent has no pro-social preferences,* $\gamma_A = 0$*, and* $\phi_P > \hat{\phi}_P$*, where* $\hat{\phi}_P$ *is defined in the appendix, then there exists a threshold* $\hat{\gamma}_P > 0$ *such that the principal delegates the control rights when* $\gamma_P \in (0, \hat{\gamma}_P)$ *and retains the control rights when* $\gamma_P > \hat{\gamma}_P$*.*

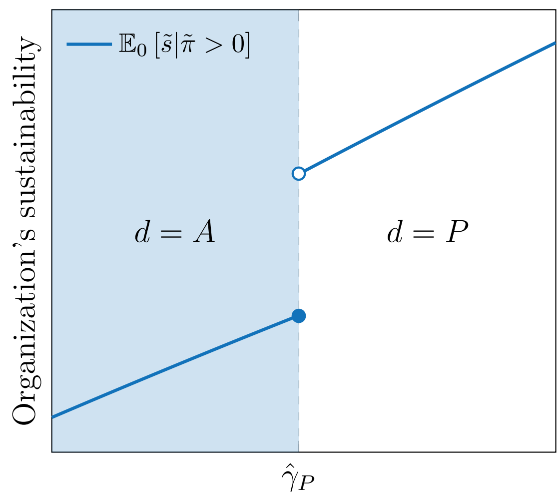


Figure 5: **Principal's pro-social preferences and organization's sustainability.** The figure plots the organization's sustainability, $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$, as a function of the principal's pro-social preferences γ_A. If $\gamma_P < \hat{\gamma}_P$, the principal delegates the control rights and retains the control rights if $\gamma_P > \hat{\gamma}_P$.

Proposition 7 shows that if the conflict of interest between the stakeholders is minor, then the principal delegates the control rights to the agent. If the conflict of interest is severe, then the cost of delegating the control rights is too high and the principal retains the control rights.

Proposition 8 (Agent without Pro-social Preferences and Organization's Sustainability)**.** *When the agent has no pro-social preferences,* $\gamma_A = 0$*, and* $\phi_P > \hat{\phi}_P$*, then an increase in the principal's pro-social preferences* γ_P *increases the organization's sustainability for all*

$\gamma_P \neq \hat{\gamma}_P$, where $\hat{\phi}_P$ and $\hat{\gamma}_P$ are the thresholds from Proposition 7, that is,

$$\forall \gamma_P \neq \hat{\gamma}_P, \quad \frac{\partial \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]}{\partial \gamma_A} \geq 0.$$

At $\hat{\gamma}_P$, the organization's sustainability increases discontinuously, that is,

$$\lim_{\gamma_P \uparrow \hat{\gamma}_P} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0] < \lim_{\gamma_P \downarrow \hat{\gamma}_P} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0].$$

Proposition 8 shows that if the agent has no pro-social preferences and the principal becomes more pro-social, then the organization's sustainability always increases, even when taking into account the change in the delegation decision. The reason is that the principal withdraws the control rights from the less pro-social agent at the threshold $\hat{\gamma}_P$, which benefits the organization's sustainability. Figure 5 shows the overall effect of increasing the principal's pro-social preferences from Proposition 8 when taking into account the principal's endogenous delegation decision.

More generally, as Proposition 9 shows even when the agent has pro-social preferences, a change in the control rights resulting from an increase in the principal's pro-social preferences benefits the organization's sustainability.

Proposition 9 (Principal's Pro-Social Preferences, Changes in Control Rights, and Organization's Sustainability)**.** *For a given set of preference parameters $\{\beta_P, \beta_A, \gamma_A\}$, assume that the delegation set is convex: $\{\gamma_P | d = A\} = [\underline{\gamma}_P, \overline{\gamma}_P]$. Further assume that the γ_P that implies no conflict of interest, $\{\gamma_P | R_P = R_A\}$, is part of the feasible parameter space.*[19] *If the allocation of the control rights changes at $\gamma'_P \in \{\underline{\gamma}_P, \overline{\gamma}_P\}$, then the organization's sustainability increases discontinuously:*

$$\lim_{\gamma_P \uparrow \gamma'_P} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0] < \lim_{\gamma_P \downarrow \gamma'_P} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0].$$

Similar to Proposition 6, the control rights can change at the boundary of the delegation set $\{\underline{\gamma}_P, \overline{\gamma}_P\}$. If they change at $\underline{\gamma}_P$, then the control rights shift from the less pro-social principal to the more pro-social agent while at $\overline{\gamma}_P$, they shift from the less pro-social agent to the more pro-social principal. In both cases, the shift in the control rights benefits the

[19] In Proposition 9, we assume that if the principal is indifferent between delegating or retaining the control rights, $\Delta U_P = 0$, then the principal delegates the control rights.

organization's sustainability.

Our results thus imply that while more pro-social principals always benefit an organization's sustainability, more pro-social agents may not. In other words, stronger pro-social preferences enhance organizational sustainability when implemented from the top-down, but may yield unintended consequences when operating from the bottom-up. In addition, our analysis highlights that even minor changes in pro-social preferences can have significant negative effects on the sustainability of an organization.

III Extensions

In this section, we study several extensions of our model. Specifically, we focus on social compensation (e.g., ESG-linked compensation) and on agent selection (e.g., hiring, internal promotions, investor selection).

A Social Compensation

In this section, we introduce the notion of social compensation into our model. While an optimal contracting approach is beyond the scope of this paper, we study how an exogenous and linear social compensation contract affects the organization's outcomes. The key insight of this section is that while social compensation can reduce the conflict of interest between stakeholders, it can also hurt the organization's sustainability through shifts in the control rights.

Even without explicitly introducing social compensation, the analysis in Section II.B already implies that paying for social performance may be misguided. That is because an organization's sustainability reflects not only the principal's choices, but also the agent's. Intuitively, if an organization's social payoff reflects the preferences and choices of multiple stakeholders, it is unclear how individual stakeholders should be rewarded for their individual choices. This broad insight poses a challenge for the design of incentive schemes for managers or employees based on ESG KPIs, which have become more prevalent. For example, as many as 57% of the S&P 500 firms currently evaluate the managers' performance based on ESG metrics (see, e.g., Ikram et al., 2019; Semler Brossy, 2021; Cohen et al., 2022; Rajan et al., 2022).

To study the additional effects of social compensation, we extend the baseline model from Section I by providing the agent with an additional monetary compensation contract that is linear in the organization's social payoff s: $\alpha_s s$. In this setting, the agent's utility from a project (π, s) is

$$u_A(\pi, s) + \alpha_s s = \beta_A \pi + (\gamma_A + \alpha_s)s.$$

The social compensation thus changes the agent's effective pro-social preferences from γ_A to $\gamma_A + \alpha_s$. We assume that $\gamma_A + \alpha_s \geq 0$. In particular, $\alpha_s = 0$ corresponds to our baseline model.

The agent's social compensation leads to a change in the agent's effective relative pro-social preferences. That is, if $\alpha_s \neq 0$, then

$$R_A^s = \frac{(\gamma_A + \alpha_s)^2}{\beta_A^2 + (\gamma_A + \alpha_s)^2} \neq \frac{\gamma_A^2}{\beta_A^2 + \gamma_A^2} = R_A,$$

and the agent's preferred project becomes $(\pi_A, s_A) = (\sqrt{1 - \iota_A}, \sqrt{\iota_A})$, where $\iota_A = R_A^s$. Consequently, the first implication of introducing social compensation is that if the principal can flexibly adjust the agent's social compensation, then it allows the principal to eliminate any conflict of interest between the two stakeholders.

Corollary 1 (Social Compensation and Project Choice)**.** *There exists a compensation contract α_s for the agent such that the effective relative pro-social preferences of the principal and agent are the same, that is, $R_P = R_A^s$, and therefore the stakeholders' preferred projects are the same, that is, $(\pi_P, s_P) = (\pi_A, s_A)$.*

While the social compensation we consider can reduce the conflict of interest in our model, there exist many constraints to using social compensation in reality. For example, if the agent is protected by limited liability, then the social compensation α_s needs to be nonnegative. In this case, social compensation can only increase the agent's effective pro-social preferences but cannot reduce them. In addition, when α_s is positive, then compensating the agent is costly for the organization. Thus, even though social compensation may be able to eliminate the conflict of interest, it may not be optimal to do so once the cost of compensation is taken into account. Social payoffs may also be hard to measure or may have multiple dimensions, which makes compensation based on these measures potentially problematic. For example, social preferences may concern issues such as the environment, social causes, or governance, which are by themselves multi-dimensional.[20]

Social compensation also impacts the agent's effort. Given that the agent's pro-social preferences effectively become $\gamma_A + \alpha_s$, the comparative statics with respect to α_s in the extended model are the same as those for γ_A in the baseline model, which we discuss in Appendix A. We can also use the results from Propositions 3 and 4 to study the effects of social compensation on the delegation decision and on the organization's sustainability. When the wedge in effective relative pro-social preferences $|R_P - R_A^s|$ is small, then the principal delegates the control rights to the agent, while when the wedge is large, then the principal retains the control rights. This implies that introducing social compensation can align preferences of the stakeholders and therefore incentivize the principal to delegate the control rights to the agent. As the following result shows, an increase in the agent's social compensation can also hurt the organization's sustainability due to the shift in the control rights.

Proposition 10 (Social Compensation, Control Rights, and Organization's Sustainability)**.** *Assume that the principal has strong relative pro-social preferences, that is, $\gamma_P > 0$ and β_P is sufficiently small. Then there exists a threshold $\hat{\alpha}_s$ such that at this threshold, the organization's sustainability decreases discontinuously, that is,*

$$\lim_{\alpha_s \uparrow \hat{\alpha}_s} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0] > \lim_{\alpha_s \downarrow \hat{\alpha}_s} \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0].$$

Figure 6 illustrates that the principal does not delegate the control rights to the agent if the conflict of interest is severe. When social compensation increases, then at some threshold the principal starts delegating the control rights to the agent. However, as the agent is still relatively less pro-social than the principal, this shift in the control rights reduces the organization's sustainability.

[20]We can show in our model that if there are multiple social payoffs and if there exists heterogeneity in preferences across stakeholders regarding the different dimensions of these payoffs, then social compensation based on a single rating measuring the overall social performance of the organization is almost surely insufficient to eliminate the conflict of interest between stakeholders.

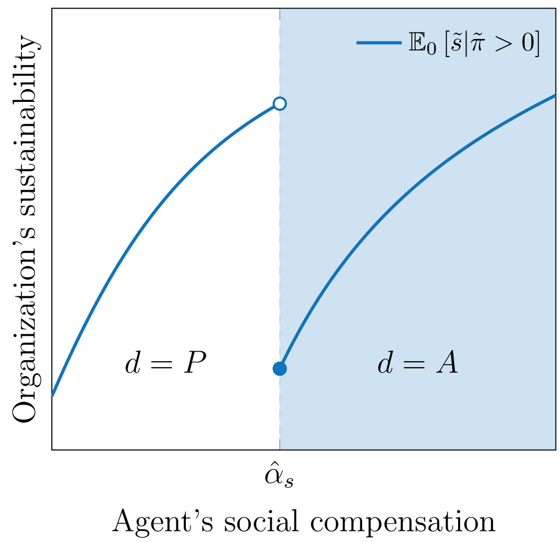


Figure 6: **Effect of agent's social compensation on organization's sustainability.** The figure plots the organization's sustainability, $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$, as a function of the agent's social compensation α_s. As the social compensation grows beyond the threshold $\hat{\alpha}_s$, the principal start delegating the control rights to the agent.

B Agent Selection

In this section, we discuss the implications of selecting an agent from a set of agents, for example by means of hiring, internal promotions or investor selection. Given that differences in relative pro-social preferences generate a conflict of interest between the principal and the agent, a crucial question to ask is whether the principal has an incentive to select an agent with different relative pro-social preferences. For example, would a manager with no pro-social preferences ever hire a pro-social employee? There exists growing empirical evidence documenting that ESG considerations play an important role for employees when selecting employers (see, e.g., Cen et al., 2022; Yao, 2022). Our analysis suggests that these considerations are also important for employers selecting employees, highlighting the other side of the matching in labor markets.

One may expect the principal to select an agent with similar pro-social preferences to avoid any conflict of interest. However, as Proposition 11 shows, this may not always be the case.

Proposition 11 (Principal without Pro-social Preferences and Agent Selection)**.** *When the principal has no pro-social preferences, $\gamma_P = 0$, and $\phi_A > \hat{\phi}_A$, where $\hat{\phi}_A$ is defined in the appendix, then there exists a threshold $\tilde{\gamma}_A$ such that selecting a marginally more pro-social* *agent improves the principal's expected utility if and only if* $\gamma_A \in (0, \tilde{\gamma}_A)$*, that is,*[21]

$$\frac{\partial \max_{d\in\{P,A\}} U_P(q_P(d), q_A(d), d)}{\partial \gamma_A} > 0 \quad \Leftrightarrow \quad \gamma_A \in (0, \tilde{\gamma}_A).$$

Furthermore, for $\gamma_A \in (0, \tilde{\gamma}_A)$*, the organization's sustainability also increases, that is, for* $\gamma_A \in (0, \tilde{\gamma}_A)$*, we have* $\frac{\partial \mathbb{E}_0[\tilde{s}|\tilde{\pi}>0]}{\partial \gamma_A} \geq 0$*.*

Keeping the monetary incentives fixed, the higher the agent's pro-social preferences are, the higher the agent's utility u_A is, independent of project choice. This, in turn, increases the agent's incentives to exert effort. Intuitively, more pro-social agents are intrinsically more motivated to exert effort because they are more concerned about the organization's sustainability. When the agent is not overly pro-social, this effort effect dominates any reduction in the principal's utility due to the diverging project choice of the agent. The principal prefers to select a pro-social agent but not an overly pro-social agent, which increases the organization's sustainability. Our model thus highlights that agent selection can benefit the sustainability of an organization.

IV Empirical Implications

In this section, we discuss the empirical implications of our main results regarding control rights, effective control, and the sustainability of organizations. Our results provide a theoretical underpinning for the increasing prevalence of stakeholder activism and engagement in addressing ESG issues.[22]

A Control Rights

Our model demonstrates that control rights can change due to increasing concerns of stakeholders regarding ESG issues and that the change in control rights can impact an organization's sustainability policy. For example, our results imply that when the conflict of interest

[21] The function $\max_{d\in\{P,A\}} U_P(q_P(d), q_A(d), d)$ is continuously differentiable except at $\hat{\gamma}_A$ from Proposition 5, where it is only continuous.

[22] See, for example, "The investor revolution" Harvard Business Review, May-June 2019, "Let employees take the lead on ESG," Wall Street Journal, June 31, 2021, "Employees demand that we become more sustainable," Forbes, October 31, 2021, and "A catalyst for greening the financial system," ECB Blog, July 8, 2022.

between the board and the CEO becomes more severe, the board of directors may withdraw control rights from the CEO. The board may do so by limiting the CEO's authority, for instance by modifying the CEO's contract or by changing corporate bylaws. In particular, a dismissal of a CEO can be interpreted as the withdrawal of all control rights. The implication regarding shifts in control rights is consistent with the findings of Huang et al. (2020), who show that disagreement between investors and management is an important driver of CEO turnover. Anecdotal evidence suggests that disagreements about pro-social policies, which correspond to the project choice in our model, can also induce turnover. For example, in 2021, Danone's CEO Emmanuel Faber was removed from his position after an attempt to transform Danone into a company that not only focuses on profits, but also on environmental sustainability.[23] An example of a manager being potentially less pro-social than his employer is illustrated by the case of HSBC Asset Management's former head of responsible investing Stuart Kirk, who was suspended after giving a controversial speech entitled "Why investors need not worry about climate risk." He left the bank shortly after.[24]

Our results related to control rights can also help explain changes in board composition and board dynamics. One measure of control rights is the number of board seats aligned with shareholders (e.g., Cotter et al., 1997; Nguyen and Nielsen, 2010). If conflicts of interests between shareholders and management arise, shareholders may initiate a proxy fight to obtain more control over the board. The engagement of the hedge fund Engine No. 1 with ExxonMobil's management serves as an example of how pro-social shareholders may attempt to gain the control rights. Engine No. 1's demands resulted in the election of three new directors to ExxonMobil's board, who challenged the company's existing strategy and pushed for a transition towards renewable energy.[25]

Another important channel through which stakeholders exercise their control rights is through shareholder proposals. Kim et al. (2019) document that firms act pro-socially at the request of their stakeholders. Consistent with this notion, the authors document that local institutional investors exert significant influence on the environmental policies of firms via shareholder proposals. Similarly, Chen et al. (2020) show that institutional shareholders use the same channel to generate improvements in social impact outcomes, while He et al.

[23]See "A top CEO was ousted after making his company more environmentally conscious. Now he's speaking out," Time Magazine, 21 November 2021.

[24]See "HSCB banker quits over climate change furore," Financial Times, 7 July 2022.

[25]See "Signs of change at ExxonMobil a year after hedge fund proxy fight," Financial Times, 24 May 2022.

(2023) provide supporting evidence based on mutual fund votes. Finally, Huang et al. (2021) provide causal evidence for this channel by documenting that a higher interest of institutional investors regarding ESG translates into more pro-social voting patterns.

B Effective Control

While formal control rights in the form of CEO contracts or board representation have important implications for sustainability outcomes, informal forms of control or soft power, which correspond to effective control in our model, also matter in the context ESG shareholder engagement. For instance, Dimson et al. (2015, 2021) document that shareholders exert informal influence on firms by persuading firms to address environmental and social issues. Hoepner et al. (2018) show that shareholder engagement on ESG issues can benefit shareholders by reducing firms' downside risks. Institutional investors play a particularly important role in shareholder engagement on ESG issues through exercising effective control. For example, Di Giuli and Kostovetsky (2014) document that political convictions of firms' stakeholders determine a firm's corporate social responsibility (CSR) spending and rating. Azar et al. (2021) provide evidence consistent with the idea that engagement by investors such as BlackRock, Vanguard, and State Street Global Advisors reduces firms' carbon emissions. This type of engagement is often informal, such as private meetings with management. One of the main channels through which this type of engagement affects corporate ESG policies is by diffusing new ESG knowledge among companies and investors (UNPRI, 2018). This diffusion of ESG knowledge is consistent with the informational friction present in our framework.

C The Sustainability of Organizations

One of the key implications of our paper is that more pro-social stakeholders can hurt the sustainability of organizations, driven by changes in control rights and effective control. An example of the negative impact of a more pro-social agent (relative to the principal) is the recent controversy surrounding BlackRock's ESG strategy. Several American states have withdrawn investment mandates from BlackRock over the concern that its incresingly important ESG strategy will have a negative impact on investor returns. Consequently, while BlackRock's strategy became more pro-social, its ability to influence the efforts of its portfolio companies to promote sustainability has been reduced.[26]

We also show that the change in control rights resulting from stakeholders becoming more pro-social can, in certain cases, increase an organization's sustainability. Importantly, our results imply that while more pro-social agents may not always benefit an organization's sustainability, more pro-social principals do. An example of the positive effect of a principal becoming more pro-social is the case of the hedge fund Engine No. 1 discussed earlier, which led to pro-social shareholders obtaining more control of ExxonMobile, allowing them to push the company to adopt a more sustainable strategy.

The ambiguous effect of more pro-social stakeholders on the sustainability of organizations is also in line with the mixed empirical evidence regarding the impact of investors on the sustainability footprint of firms and other organizations. The potentially negative effect is documented by Kim et al. (2022), who show that issuing ESG-linked loans can lead to a deterioration in ESG scores. Hartzmark and Shue (2023) highlight that sustainable investing practices may be counterproductive by making brown firms more brown without improving the sustainability of green firms. Heath et al. (2021) show that there is no effect, by demonstrating that socially responsible investment funds do not improve the behavior of their portfolio companies in terms of environmental and social outcomes. In contrast, other studies document a positive effect of stakeholders on firms' sustainability (e.g., Di Giuli and Kostovetsky, 2014; Chen et al., 2020; Huang et al., 2021; Gantchev et al., 2022). Our analysis thus highlights that when assessing the impact of pro-social stakeholders, it is crucial to assess whether it is associated with a change in control rights and whether the change in control rights is driven by more pro-social principals or agents, as this informs whether the impact on a firm's sustainability is positive or negative.

V Conclusion

We develop a theory of stakeholder governance to study how stakeholders with pro-social preferences influence an organization's sustainability. Our analysis highlights that more pro-social stakeholders may not always improve the organization's sustainability. That happens

[26]See "Florida to pull $2bn from BlackRock in spreading ESG backlash," Financial Times, December 1, 2022.

because the control rights may shift from the more- to the less pro-social stakeholder due to conflicts of interest over preferred sustainability policies. In addition to analyzing how pro-social stakeholders impact control rights, effective control and organizations' sustainability, we also study how ESG compensation and agent selection (e.g., hiring, investor selection, internal promotions) affect organizations. We find that social compensation can negatively impact an organization's sustainability and that organizations prefer to select more but not overly pro-social agents.

Our analysis can be applied to different types of relationships between stakeholders such as managers and employees, entrepreneurs and investors, company boards and CEOs, as well as regulators and firms. In general, our model provides a theoretical underpinning for the increasing prevalence of ESG-related stakeholder activism and engagement.

References

Aghion, P. and Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy*, 105(1):1–29.

Allen, F., Carletti, E., and Marquez, R. (2015). Stakeholder governance, competition, and firm value. *Review of Finance*, 19(3):1315–1346.

Alonso, R., Dessein, W., and Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, 98(1):145–79.

Azar, J., Duro, M., Kadach, I., and Ormazabal, G. (2021). The big three and corporate carbon emissions around the world. *Journal of Financial Economics*, 142(2):674–696.

Baker, G., Gibbons, R., and Murphy, K. J. (1999). Informal authority in organizations. *Journal of Law, Economics, and Organization*, 15(1):56–73.

Bolton, P. and Dewatripont, M. (2013). Authority in organizations. *Handbook of Organizational Economics*, pages 342–372.

Bonnefon, J.-F., Landier, A., Sastry, P., and Thesmar, D. (2019). Do investors care about corporate externalities? Experimental evidence. *Working Paper, HEC Paris.*

Broccardo, E., Hart, O., and Zingales, L. (2022). Exit vs. voice. *Forthcoming, Journal of Political Economy.*

Burkart, M., Gromb, D., and Panunzi, F. (1997). Large shareholders, monitoring, and the value of the firm. *Quarterly Journal of Economics*, 112(3):693–728.

Cen, X., Qiu, Y., and Yue Wang, T. (2022). Corporate social responsibility and employee retention. *Working Paper, Texas A&M University.*

Chen, J. (2022). Optimal managerial authority. *Working Paper, Boston University.*

Chen, T., Dong, H., and Lin, C. (2020). Institutional shareholders and corporate social responsibility. *Journal of Financial Economics*, 135(2):483–504.

Chowdhry, B., Davies, S. W., and Waters, B. (2019). Investing for impact. *Review of Financial Studies*, 32(3):864–904.

Cohen, S., Kadach, I., Ormazabal, G., and Reichelstein, S. (2022). Executive compensation tied to ESG performance: International evidence. *Working Paper, Stanford University.*

Cotter, J. F., Shivdasani, A., and Zenner, M. (1997). Do independent directors enhance target shareholder wealth during tender offers? *Journal of Financial Economics*, 43(2):195–218.

Dasgupta, A., Fos, V., and Sautner, Z. (2021). Institutional investors and corporate governance. *Foundations and Trends in Finance*, 12(4):276–394.

Dessein, W. (2002). Authority and communication in organizations. *Review of Economic Studies*, 69(4):811–838.

Di Giuli, A. and Kostovetsky, L. (2014). Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of Financial Economics*, 111(1):158–180.

Dimson, E., Karakaş, O., and Li, X. (2015). Active ownership. *Review of Financial Studies*, 28(12):3225–3268.

Dimson, E., Karakaş, O., and Li, X. (2021). Coordinated engagements. *Working Paper, University of Cambridge*.

Dyck, A., Lins, K. V., Roth, L., and Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. *Journal of Financial Economics*, 131(3):693–714.

Eccles, R. G. and Klimenko, S. (2019). The investor revolution. *Harvard Business Review*.

Gantchev, N., Giannetti, M., and Li, R. (2022). Does money talk? divestitures and corporate environmental and social policies. *Review of Finance*, 26(6):1469–1508.

Gibson, R., Krueger, P., and Mitali, S. (2021). The sustainability footprint of institutional investors: ESG driven price pressure and performance. *Working Paper, University of Geneva*.

Gillan, S. L., Koch, A., and Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, 66:101889.

Gollier, C. and Pouget, S. (2022). Investment strategies and corporate behaviour with socially responsible investors: A theory of active ownership. *Economica*, 89(356):997–1023.

Green, D. and Roth, B. (2021). The allocation of socially responsible capital. *Working Paper, Harvard Business School*.

Grenadier, S. R., Malenko, A., and Malenko, N. (2016). Timing decisions in organizations: Communication and authority in a dynamic environment. *American Economic Review*, 106(9):2552–81.

Gupta, D., Kopytov, A., and Starmans, J. (2022). The pace of change: Socially responsible investing in private markets. *Working Paper, Stockholm School of Economics*.

Harris, M. and Raviv, A. (2005). Allocation of decision-making authority. *Review of Finance*, 9(3):353–383.

Hart, O. and Zingales, L. (2022). The new corporate governance. *Working Paper, University of Chicago*.

Hartzmark, S. M. and Shue, K. (2023). Counterproductive sustainable investing: The impact elasticity of brown and green firms. *Working Paper, Yale School of Management*.

He, Y., Kahraman, B., and Lowry, M. (2023). ES risks and shareholder voice. *Forthcoming, Review of Financial Studies*.

Heath, D., Macciocchi, D., Michaely, R., and Ringgenberg, M. C. (2021). Does socially responsible investing change firm behavior? *Forthcoming, Review of Finance*.

Heeb, F., Kölbel, J. F., Paetzold, F., and Zeisberger, S. (2022). Do investors care about impact? *Forthcoming, Review of Financial Studies*.

Heinkel, R., Kraus, A., and Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, 36(4):431–449.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., and Zhou, X. (2018). ESG shareholder engagement and downside risk. *Working Paper, University of Texas at Austin*.

Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1):15–36.

Hong, H. and Kostovetsky, L. (2012). Red and blue investing: Values and finance. *Journal of Financial Economics*, 103(1):1–19.

Huang, S., Maharjan, J., and Thakor, A. V. (2020). Disagreement-induced CEO turnover. *Journal of Financial Intermediation*, 43:100819.

Huang, W., Karolyi, G. A., and Kwan, A. (2021). Paying attention to ESG matters: Evidence from big data analytics. *Working Paper, Hong Kong University Business School*.

Ikram, A., Li, Z. F., and Minor, D. (2019). CSR-contingent executive compensation contracts. *Forthcoming, Journal of Banking and Finance*.

Kim, I., Wan, H., Wang, B., and Yang, T. (2019). Institutional investors and corporate environmental, social, and governance policies: Evidence from toxics release data. *Management Science*, 65(10):4901–4926.

Kim, S., Kumar, N., Lee, J., and Oh, J. (2022). ESG lending. *Working Paper, University of Florida*.

Krueger, P., Metzger, D., and Wu, J. (2022). The sustainability wage gap. *Working Paper, University of Geneva*.

Krueger, P., Sautner, Z., and Starks, L. T. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*, 33(3):1067–1111.

Landier, A. and Lovo, S. (2020). ESG investing: How to optimize impact? *Working Paper, HEC Paris*.

Levit, D., Malenko, N., and Maug, E. G. (2022). Trading and shareholder democracy. *Forthcoming, Journal of Finance*.

List, J. A. (2009). Social preferences: Some thoughts from the field. *Annual Review of Economics*, 1(1):21–1.

Magill, M., Quinzii, M., and Rochet, J.-C. (2015). A theory of the stakeholder corporation. *Econometrica*, 83(5):1685–1725.

Malenko, N. (2022). Information flows, organizational structure, and corporate governance. *Working Paper, University of Michigan*.

Matsusaka, J. G. and Shu, C. (2021). A theory of proxy advice when investors have social goals. *Working Paper, University of Southern California*.

McCahery, J. A., Sautner, Z., and Starks, L. T. (2016). Behind the scenes: The corporate governance preferences of institutional investors. *Journal of Finance*, 71(6):2905–2932.

Morgan, J. and Tumlinson, J. (2019). Corporate provision of public goods. *Management Science*, 65(10):4489–4504.

Nguyen, B. D. and Nielsen, K. M. (2010). The value of independent directors: Evidence from sudden deaths. *Journal of Financial Economics*, 98(3):550–567.

Oehmke, M. and Opp, M. M. (2022). A theory of socially responsible investment. *Working Paper, Stockholm School of Economics*.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.

Pedersen, L. H. and Feldhütter, P. (2022). Is capital structure irrelevant with ESG investors? *Working Paper, Copenhagen Business School*.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.

Rajan, R. G., Ramella, P., and Zingales, L. (2022). What purpose do corporations purport? Evidence from letters to shareholders. *Working Paper, University of Chicago*.

Riedl, A. and Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *Journal of Finance*, 72(6):2505–2550.

Roth, B. (2021). Impact investing: A theory of financing social enterprises. *Working Paper, Harvard Business School*.

Semler Brossy (2021). ESG+Incentives 2021 report.

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance*, 57(5):1891–1921.

Tirole, J. (2001). Corporate governance. *Econometrica*, 69(1):1–35.

UNPRI (2018). How ESG engagement creates value for investors and companies. *United Nations Principles for Responsible Investing*.

Van den Steen, E. (2010a). Culture clash: The costs and benefits of homogeneity. *Management Science*, 56(10):1718–1738.

Van den Steen, E. (2010b). On the origin of shared beliefs (and corporate culture). *The RAND Journal of Economics*, 41(4):617–648.

Yao, T. (2022). ESG education and job choice. *Working Paper, HEC Paris.*

Appendix

The first part of the appendix discusses the impact of monetary incentives and pro-social preferences on effort and effective control. The second part shows that our results do not depend on the discrete nature of the allocation of control rights. The third part contains the proofs.

A Effort and Effective Control

This appendix studies how monetary incentives and pro-social preferences affect stakeholders' incentives to exert effort and their allocation of effective control.

As the following result shows, changes in the stakeholders' incentive and preference parameters lead to a substitution between their effort levels in equilibrium.

Proposition 12 (Monetary Incentives, Pro-social Preferences, and Effort)**.** *A change in the principal's or agent's monetary incentives or pro-social preferences leads to a substitution of effort between the principal and the agent. That is, for* $\theta \in \{\beta_P, \gamma_P, \beta_A, \gamma_A\}$*,*

$$\frac{\partial q_P(d)}{\partial \theta}\frac{\partial q_A(d)}{\partial \theta} \leq 0.$$

The substitution of effort between the principal and agent, as demonstrated in Proposition 12, implies a substitution of effective control between the two stakeholders. This substitution has the same direction as the substitution of effort.

Corollary 2 (Monetary Incentives, Pro-social Preferences, and Effective Control)**.** *A change in the principal's or agent's monetary incentives or pro-social preferences leads to a substitution of effective control between the principal and the agent in line with their changes in effort. That is, for* $\theta \in \{\beta_P, \gamma_P, \beta_A, \gamma_A\}$*,*

$$\frac{\partial e_P(d)}{\partial \theta}\frac{\partial e_A(d)}{\partial \theta} \leq 0 \quad \textit{and} \quad \frac{\partial e_P(d)}{\partial \theta}\frac{\partial q_P(d)}{\partial \theta} \geq 0 \quad \textit{and} \quad \frac{\partial e_A(d)}{\partial \theta}\frac{\partial q_A(d)}{\partial \theta} \geq 0.$$

Proposition 12 and Corollary 2 highlight a key force that arises in our model. In addition to changing the preferred project, altering a stakeholder's monetary incentives or pro-social preferences also leads to a substitution of effort from one stakeholder to the other and thus to a substitution of effective control. In particular, making a stakeholder more pro-social not only shifts the stakeholder's preferred project towards the social payoff, but also changes the allocation of effective control—the extent to which the stakeholder can actually influence the project the organization eventually implements.

To understand the effect a more pro-social stakeholder has on the organization's payoffs, we need to analyze how altering pro-social preferences affects effective control. For instance, do more pro-social employees have a larger or a smaller impact on the firm? To this end, we first study the effect of changing the agent's pro-social preferences γ_A. We then discuss the comparative statics with respect to the principal's pro-social preferences γ_P as well as the stakeholders' monetary incentives β_P and β_A.[27]

Proposition 13 (Pro-social Preferences and Effort when Principal Holds Control Rights)**.** *When the principal holds the control rights, $d = P$, then the principal exerts less effort and the agent exerts more effort when the agent's pro-social preferences γ_A increase, that is,*

$$\frac{\partial q_P(P)}{\partial \gamma_A} \leq 0 \quad \textit{and} \quad \frac{\partial q_A(P)}{\partial \gamma_A} \geq 0.$$

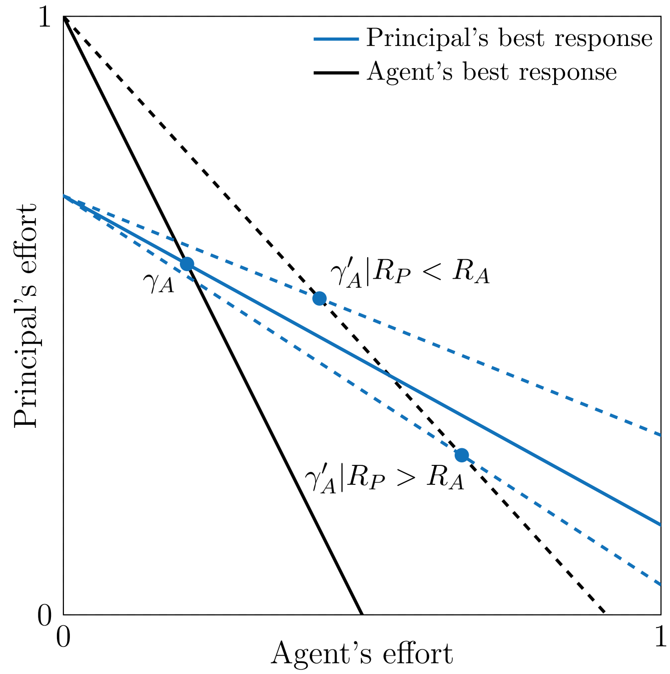


Figure A.1: **Agent's pro-social preferences and equilibrium effort when the principal holds the control rights.** The figure plots the principal's best response function $B_P(q_A, P)$ and the agent's best response function $B_A(q_P, P)$. The solid lines are the best response functions for some initial level of the agent's pro-social preferences γ_A and the dashed lines for a marginally higher level of the agent's pro-social preferences $\gamma_A' > \gamma_A$. The figure distinguishes between two cases of the principal's best response function, one in which the initial γ_A satisfies $R_P > R_A$ and one in which it satisfies $R_P < R_A$.

Strengthening the agent's pro-social preferences unambiguously increases the agent's incentives to exert effort. Figure A.1 illustrates this by showing that the agent's best response function $B_A(q_A, P)$ shifts outwards as γ_A increases. This happens because the agent's utility when having effective control, $u_A(\pi_A, s_A)$, increases. Intuitively, because the agent cares

[27]Note that the comparative statics with respect to the principal's and agent's effort costs can be directly determined from their impact on the best response functions. A higher effort cost ϕ_P of the principal decreases the incentives to exert effort and therefore lowers the principal's best response function, while the best response function of the agent remains unaffected. Therefore, the principal's effort decreases while the agent's effort increases. A similar argument can be made when increasing the agent's effort cost ϕ_A.

more about the organization's social payoff, the utility when having effective control increases, which in turn increases the incentives to exert effort.

Importantly, the principal's best response function and therefore incentives to exert effort are also affected by a change in the agent's pro-social preferences, that is,

$$\frac{\partial B_P(q_A, P)}{\partial \gamma_A} = -q_A \frac{\frac{\partial u_P(\pi_A, s_A)}{\partial \gamma_A}}{\phi_P} < 0 \quad \Leftrightarrow \quad R_P > R_A.$$

This happens because the agent's project choice, and therefore the principal's utility, changes when the agent has effective control. If the principal has stronger relative pro-social preferences than the agent, that is, if $R_P > R_A$, then the principal's utility when the agent's preferred project (π_A, s_A) is implemented increases in response to a higher γ_A as it brings the agent's preferred project closer to the principal's. Thus, losing effective control to the agent becomes less costly, which in turn reduces the principal's incentives to exert effort. Put differently, the agent's effective control provides a better hedge for the principal in this case. In contrast, if an increase in the agent's pro-social preferences aggravates the conflict of interest between the stakeholders, that is, if $R_P < R_A$, the principal's incentives to exert effort increase.

The two cases are illustrated in Figure A.1. It turns out that in our model, the direct effect on the agent's utility always dominates the indirect effect on the principal's utility. As such, a higher γ_A increases the agent's effort and decreases the principal's effort in equilibrium. For example, even if the manager of a firm controls the firm's decision making, employees becoming more pro-social causes them to gain more effective control and therefore more influence on the firm's outcomes.

Proposition 14 (Pro-social Preferences and Effort when Agent Holds Control Rights)**.** *When the agent holds the control rights, $d = A$, then there exists a threshold $\tilde{\gamma}_A$ such that when $\gamma_A < \tilde{\gamma}_A$, an increase in the agent's pro-social preferences increases the principal's effort and decreases the agent's effort, that is,*

$$\frac{\partial q_P(A)}{\partial \gamma_A} \geq 0 \quad \textit{and} \quad \frac{\partial q_A(A)}{\partial \gamma_A} \leq 0,$$

and vice verse when $\gamma_A > \tilde{\gamma}_A$, that is,

$$\frac{\partial q_P(A)}{\partial \gamma_A} \leq 0 \quad \textit{and} \quad \frac{\partial q_A(A)}{\partial \gamma_A} \geq 0.$$

Proposition 14 highlights that the effect of strengthening a stakeholder's pro-social preferences on the allocation of effective control critically depends on whether that stakeholder holds the control rights. As is clear from the principal's best response function, $B_P(q_A, A)$, the incentives to exert effort do not directly depend on the agent's preferred project (π_A, s_A) and therefore on γ_A. Intuitively, because the agent holds the control rights, the principal cannot directly reduce the probability that the agent has effective control and therefore the agent's preferred project does not directly affect the principal's effort incentives.

In contrast, the agent's best response function, and therefore incentives to exert effort, depends on pro-social preferences γ_A. In particular, we have

$$\frac{\partial B_A(q_P, A)}{\partial \gamma_A} = (1-q_P)\frac{\frac{\partial u_A(\pi_A, s_A)}{\partial \gamma_A}}{\phi_A} + q_P\frac{\frac{\partial \Delta u_A}{\partial \gamma_A}}{\phi_A} = \underbrace{(1-q_P)\frac{s_A}{\phi_A}}_{\textit{Direct Utility Effect}} + \underbrace{q_P\frac{(s_A - s_P)}{\phi_A}}_{\textit{Hedging Effect}}.$$

The first term captures the direct effect on the agent's utility, which is always positive because increasing the agent's pro-social preferences results in a higher utility level. The second—hedging—effect arises as the agent can delegate project choice to the principal if failing to generate information. If γ_A is sufficiently low, such that $R_A < R_P \Leftrightarrow s_A < s_P$, the hedging effect is negative because the conflict of interest between the stakeholders becomes less severe as γ_A increases, which makes the hedge more valuable and therefore lowers the agent's incentives to exert effort. On the other hand, if γ_A is sufficiently high, such that $R_A > R_P \Leftrightarrow s_A > s_P$, the hedging effect is positive. Taken together, when γ_A is low, the hedging effect dominates and therefore the agent's effort decreases and the principal's increases as the agent becomes more pro-social. At $\gamma_A = \tilde{\gamma}_A$, the direct utility effect starts to dominate.[28] A further increase in γ_A thus leads to the agent exerting more effort and to the principal exerting less effort.

The comparative statics with respect to the principal's pro-social preferences γ_P as well as the stakeholders' monetary incentives β_P and β_A follow from the results obtained above due to the symmetry of our model. Specifically, we can relabel the social payoff as the monetary payoff and therefore the comparative statics with respect to β_A and γ_A are qualitatively identical. Furthermore, we can interchange the role of the principal and the agent conditional on the delegation decision. As a result, the comparative statics with respect to γ_A and β_A when $d = A$ ($d = P$) are qualitatively identical to those with respect to γ_P and β_P when $d = P$ ($d = A$).

B Continuous Delegation of Control Rights

In this section, we show that our result in Proposition 5 that an increase in the agent's pro-social preferences can reduce the organization's sustainability due to a withdrawal of the control rights does not rely on the fact that the change in the control rights is discrete. As we show below in a simple extension of our model, what we need is that strengthening the agent's pro-social preferences leads to a significant reduction in the agent's control rights.

In the baseline model, the organization has a single task, namely it needs to decide which project to undertake, if any at all. Assume now that the organization has $N > 1$ tasks indexed by $i \in \{1, ..., N\}$ instead of a single one. Each task i consists of a project choice similar to the one in the baseline model. The payoffs from task i are $\frac{1}{N}$ times the payoffs of a project from the baseline model and zero if no project is undertaken, that is, the payoffs for task i from the principal's and agent's preferred projects are given by $\frac{1}{N}(\pi_P, s_P)$ and

[28] The condition $(1 - q_P(A))s_A + q_P(A)(s_A - s_P) = 0$ implicitly characterizes $\tilde{\gamma}_A$ when it is positive. Observe that at $R_P = R_A$, the hedging effect turns from negative to positive. This in combination with the fact that the direct utility effect always gives the agent stronger incentives to exert effort implies that the threshold $\tilde{\gamma}_A$ must be below the level at which $R_P = R_A$. Thus, the threshold $\tilde{\gamma}_A$ satisfies $R_A < R_P$.

$\frac{1}{N}(\pi_A, s_A)$, respectively, and the effort cost to learn about the project payoffs for task i is $\frac{1}{N}$ times the effort cost in the baseline model, that is, $\frac{1}{N}\frac{\phi_j}{2}q_j^2$, where $j \in \{P, A\}$. In addition, the principal receives an extra utility $\frac{\epsilon_i}{N}$ from retaining the control rights for task i, where the random variables ϵ_i, $i \in \{1, ..., N\}$, are independently drawn from a uniform distribution with support $[-\sigma, \sigma]$ with $\sigma \geq 0$. This setup implies that for each task i, the principal delegates the control rights to the agent when $\Delta U_P^i = \frac{1}{N}(\Delta U_P + \epsilon_i) < 0$ and retains the control rights when $\Delta U_P^i > 0$.

As a result, when $N \to \infty$, because of the law of large numbers, the fraction of tasks for which the principal delegates control rights to the agent is one when $\Delta U_P < -\sigma$, $\frac{\sigma - \Delta U_P}{2\sigma}$ when $\Delta U_P \in [-\sigma, \sigma]$, and zero when $\Delta U_P > \sigma$. If $\sigma = 0$, then we are back to our baseline model. For $\sigma > 0$, the fraction of the control rights delegated to the agent changes continuously as γ_A changes.

For N finite, the organization's sustainability is the sum over the tasks $i \in \{1, ..., N\}$ of the expected social payoff conditional on a project being undertaken for that task. Observe that when $\sigma > 0$ and $N \to \infty$, the organization's sustainability is continuous in γ_A. Furthermore, when σ gets sufficiently small, there exist two thresholds γ_A' and γ_A'' satisfying $\gamma_A' < \hat{\gamma}_A < \gamma_A''$, where $\hat{\gamma}_A$ is defined in Proposition 5, such that:

i) For γ_A', the principal delegates authority for all tasks while for γ_A'' the principal retains authority for all tasks. Thus, for γ_A' and γ_A'', the organization's sustainability is the same as in the baseline model.

ii) For γ_A' and γ_A'', the organization's sustainability, which is the same as in the baseline model, satisfies

$$\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0, \gamma_A'] > \mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0, \gamma_A''].$$

Continuity of the organization's sustainability in γ_A then implies that there exists a $\gamma_A \in [\gamma_A', \gamma_A'']$ such that $\frac{\partial \mathbb{E}_0[\tilde{s}|\tilde{\pi}>0]}{\partial \gamma_A} < 0$.

C Proofs

We organize the proofs into three sections. The first section contains the proofs for the baseline model related to the effort results taking the delegation of the control rights as given (Section II.B and Appendix A). The second section contains the proofs for the baseline model related to the delegation of the control rights (Section II.C). The third section contains the proofs for the model extensions (Section III).

I Proofs for Section II.B and Appendix A

Lemma 3 (Equilibrium Effort Choices)**.** *Given the delegation decision d, there exists a unique Nash equilibrium in effort choices $(q_P(d), q_A(d)) \in (0, 1)^2$ at time one, which is the solution to the first-order conditions of the principal's expected utility and the agent's expected utility with respect to their effort levels.*

Proof of Lemma 3. First, we want to show that the lower bounds for ϕ_P and ϕ_A ensure that $q_P < 1$ and $q_A < 1$. If $q_P = q_A = 1$ then the agent without the control rights would be better off setting the effort to zero. Therefore, either $q_P < 1$ or $q_A < 1$. Assume, without loss of generality, that $q_P < 1$, then the lower bound for ϕ_A implies that $q_A < 1$. Therefore, $q_P < 1$ and $q_A < 1$. Second, from the first-order conditions it directly follows that $q_A > 0$ and $q_P > 0$. As a consequence, the two first-order conditions define the optimal effort levels.

Finally, the first-order conditions define a system of two linear equations with two unknowns (q_P, q_A). Direct calculations allow us to show that this system has a unique solution $(q_P(d), q_A(d))$. □

Proof of Proposition 12. Given $d = P$, we need to sign the product of four different pairs of derivatives. We start by signing two after which the other two follow from symmetry within the model.

1. Observe that

$$\frac{\partial q_P(P)}{\partial \gamma_A} = \frac{\gamma_P \phi_A \left(\sqrt{\beta_P^2 + \gamma_P^2} - \phi_P\right)}{(\beta_A\beta_P + \gamma_A\gamma_P - \phi_A\phi_P)^2} \leq 0,$$
$$\frac{\partial q_A(P)}{\partial \gamma_A} = \frac{\left(\sqrt{\beta_P^2 + \gamma_P^2} - \phi_P\right)(\beta_A(\beta_P\gamma_A - \beta_A\gamma_P) - \gamma_A\phi_A\phi_P)}{\sqrt{\beta_A^2 + \gamma_A^2}(\beta_A\beta_P + \gamma_A\gamma_P - \phi_A\phi_P)^2} \geq 0.$$

The first inequality follows from

$$\begin{aligned}(\beta_A\beta_P + \gamma_A\gamma_P) &= u_P(\pi_P, s_P)u_A(\pi_P, s_P) \\ &\leq u_P(\pi_P, s_P)u_A(\pi_A, s_A) \\ &= \sqrt{\beta_P^2 + \gamma_P^2}\sqrt{\beta_A^2 + \gamma_A^2} \\ &< \phi_P\phi_A. \end{aligned} \tag{A.1}$$

and the fact that

$$\sqrt{\beta_P^2 + \gamma_P^2} = u_P(\pi_P, s_P) < \phi_P. \tag{A.2}$$

While the second inequality follows from equation (A.1), equation (A.2), and the fact that

$$\begin{aligned}\beta_A(\beta_P\gamma_A - \beta_A\gamma_P) - \gamma_A\phi_A\phi_P &\leq \beta_A(\beta_P\gamma_A - \beta_A\gamma_P) - \gamma_A(\beta_A\beta_P + \gamma_A\gamma_P) \\ &= -(\beta_A^2 + \gamma_A^2)\gamma_P \leq 0.\end{aligned}$$

2. Observe that

$$\frac{\partial q_P(P)}{\partial \gamma_P}\frac{\partial q_A(P)}{\partial \gamma_P} =$$
$$-\frac{\phi_A\sqrt{\beta_A^2 + \gamma_A^2}\left(-\beta_A\beta_P\gamma_P - \gamma_A\phi_P\sqrt{\beta_P^2 + \gamma_P^2} + \beta_P^2\gamma_A + \gamma_P\phi_A\phi_P\right)^2}{(\beta_P^2 + \gamma_P^2)(\beta_A\beta_P + \gamma_A\gamma_P - \phi_A\phi_P)^4} \leq 0.$$

The denominator is positive because of equation (A.1), which proves the inequality.

The comparative statics with respect stakeholders' monetary incentives β_A and β_P follow from the results obtained above due to the symmetry in our model. Specifically, we can relabel the social payoff as the monetary payoff and therefore the comparative statics with respect to β_A (β_P) and γ_A (γ_P) are qualitatively identical.

Furthermore, we can interchange the role of the principal and the agent conditional on the delegation decision to obtain the results when $d = A$. □

Proof of Corollary 2. The first result follows directly from that fact that $e_P(d) = 1 - e_A(d)$ and therefore

$$\frac{\partial e_P(d)}{\partial \theta}\frac{\partial e_A(d)}{\partial \theta} \leq 0.$$

Observe that

$$e_P^{-1}(P) = \frac{q_P(P) + (1 - q_P(P))q_A(P)}{q_P(P)} = 1 + \frac{1 - q_P(P)}{q_P(P)}q_A(P).$$

If $q_P(P)$ increases then $q_A(P)$ decreases and therefore $\frac{1-q_P(P)}{q_P(P)}q_A(P)$ decreases and $e_P(P)$ increases. Therefore,

$$\frac{\partial e_P(P)}{\partial \theta}\frac{\partial q_P(P)}{\partial \theta} \geq 0.$$

Similar arguments show that

$$\frac{\partial e_P(A)}{\partial \theta}\frac{\partial q_P(A)}{\partial \theta} \geq 0.$$

The final result follows from the symmetry in the model and can be obtained by interchanging the principal and the agent in the steps above. □

Proof of Proposition 13. The result follows directly from the derivations in the proof of Proposition 12. □

Proof of Proposition 14. Given that the agent holds the control rights, $d = A$, the principal's best response function $B_P(q_A, A)$ does not change when varying γ_A. The agent's best response function is given by

$$B_A(q_P, A) = \frac{(1 - q_P)u_A(\pi_A, s_A) + q_P\Delta u_A}{\phi_A}.$$

From the envelope theorem it then follows that

$$\frac{\partial B_A(q_P, A)}{\partial \gamma_A} = \frac{s_A - q_P s_P}{\phi_A}.$$

Therefore, if $s_A - q_P s_P > 0$, then $\frac{\partial q_A(A)}{\partial \gamma_A} > 0$ and, as implied by Proposition 12, $\frac{\partial q_P(A)}{\partial \gamma_A} \leq 0$. Similarly, if $s_A - q_P s_P < 0$, then $\frac{\partial q_A(A)}{\partial \gamma_A} < 0$ and, as implied by Proposition 12, $\frac{\partial q_P(A)}{\partial \gamma_A} \geq 0$. Finally, if $s_A - q_P s_P = 0$, then $\frac{\partial B_A(q_P,A)}{\partial \gamma_A} = \frac{\partial B_P(q_A,A)}{\partial \gamma_A} = 0$, and therefore $\frac{\partial q_A(A)}{\partial \gamma_A} = \frac{\partial q_P(A)}{\partial \gamma_A} = 0$.

Assume that $s_A - q_P s_P \geq 0$, then s_A is increasing in γ_A while q_P is weakly decreasing in γ_A and s_P remains unchanged. Therefore, if there exists a γ_A such that $s_A - q_P s_P \geq 0$, then $s_A - q_P s_P > 0$ for any $\gamma'_A > \gamma_A$. Furthermore, we know that when $R_A > R_P$, then $s_A - q_P s_P > s_A - s_P > 0$. Taken together, this proves that there exists a unique $\tilde{\gamma}_A$ such that $s_A - q_P s_P = 0$. For $\gamma'_A > \tilde{\gamma}_A$, we have $s_A - q_P s_P > 0$, and for $\gamma'_A < \tilde{\gamma}_A$, we have $s_A - q_P s_P < 0$, which proves the result. □

Proof of Lemma 1. When $d = P$, the result follows directly from Proposition 13 and Corollary 2.

When $d = A$ and $\gamma_P = 0$, then $\tilde{\gamma}_A = 0$ because

$$s_A - q_P s_P = s_A \geq 0.$$

The result then follows from Proposition 14 and Corollary 2. □

Proof of Proposition 1. The result follows from Lemma 1 and the fact that

$$\mathbb{E}_0 \left[\tilde{s} | \tilde{\pi} > 0, d \right] = e_A(d) s_A.$$

□

II Proofs for Section II.C

Proof of Lemma 2. Direct calculations imply that

$$q_P(P) - q_P(A) = q_A(A) - q_A(P) = \frac{-\sqrt{\left(\beta_A^2 + \gamma_A^2\right)\left(\beta_P^2 + \gamma_P^2\right)} + \beta_A\beta_P + \gamma_A\gamma_P}{\beta_A\beta_P + \gamma_A\gamma_P - \phi_A\phi_P} \geq 0.$$

The inequality follows from equation (A.1), which shows that the numerator and denominator are nonpositive and negative, respectively. □

Proof of Proposition 2. Since $R_P = R_A$, it follows that $(\pi_P, s_P) = (\pi_A, s_A)$ and therefore $\Delta u_P = \Delta u_A = 0$. The first-order conditions that determine the effort levels are therefore independent of d and as a result

$$(q_P(P), q_A(P)) = (q_P(A), q_A(A)).$$

These observations then imply that

$$U_P(q_P(P), q_A(P), P) = U_P(q_P(A), q_A(A), A),$$

which completes the proof. □

Proof of Proposition 3. We first prove the result for when the wedge is small after which we prove the result for when the wedge is large.

Take a set of parameters $(\beta'_P, \gamma'_P, \beta'_A, \gamma'_A)$. Define $(\beta'_P, \gamma'_P, \beta'_A, \gamma''_A)$, where $\gamma''_A = \beta'_A\gamma'_P/\beta'_P \geq 0$. For this set of parameters, the relative pro-social preferences are the same and from

Proposition 2 it then follows that $\Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A'') = 0$. As a consequence,

$$\Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A') = \int_{\gamma_A''}^{\gamma_A'} \frac{\partial \Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A)}{\partial \gamma_A} d\gamma_A.$$

Observe that

$$\left.\frac{\partial \Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A)}{\partial \gamma_A}\right|_{\gamma_A=\gamma_A''} = 0,$$

$$\left.\frac{\partial^2 \Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A)}{\partial \gamma_A^2}\right|_{\gamma_A=\gamma_A''} = \frac{(\beta_P')^4\left(\left(\beta_P'\right)^2 + (\gamma_P')^2 - \sqrt{\left(\beta_P'\right)^2 + (\gamma_P')^2}\phi_P\right)}{\sqrt{\left(\beta_P'\right)^2 + (\gamma_P')^2}\left(\beta_A'\left(\left(\beta_P'\right)^2 + (\gamma_P')^2\right) - \beta_P'\phi_A\phi_P\right)^2} < 0,$$

because of equation (A.2) and the fact that when the relative pro-social preferences are aligned (i.e., $R_P = R_A$), then

$$u_P(\pi_P, s_P)u_A(\pi_P, s_P) = \frac{\beta_A}{\beta_P}(\beta_P^2 + \gamma_P^2) < \phi_A\phi_P.$$

Note that

$$|R_P' - R_A'| = \left|\frac{\left(\frac{\gamma_P'}{\beta_P'}\right)^2}{\left(\frac{\gamma_P'}{\beta_P'}\right)^2 + 1} - \frac{\left(\frac{\gamma_A'}{\beta_A'}\right)^2}{\left(\frac{\gamma_A'}{\beta_A'}\right)^2 + 1}\right| = \left|\frac{(r_P')^2}{(r_P')^2 + 1} - \frac{(r_A')^2}{(r_A')^2 + 1}\right|,$$

where $r_P' := \frac{\gamma_P'}{\beta_P'}$ and $r_A' := \frac{\gamma_A'}{\beta_A'}$, and

$$|\gamma_A'' - \gamma_A'| = \left|\frac{\beta_A'}{\beta_P'}\gamma_P' - \gamma_A'\right| = \beta_A'\left|r_P' - r_A'\right| \leq \phi_A\left|r_P' - r_A'\right|,$$

because $\beta_A' \leq \sqrt{(\beta_A')^2 + (\gamma_A')^2} < \phi_A$. Thus, for given parameters γ_P' and β_P', as $|R_P' - R_A'| \to 0$, we have that $|r_P' - r_A'| \to 0$ because the function $f(r) = \frac{r^2}{r^2+1}$ is continuous and strictly increasing for $r > 0$. Thus, if $|R_P' - R_A'|$ is sufficiently small, then γ_A' is sufficiently close to γ_A''.

Continuity of ΔU_P and its first- and second-order derivative with respect to γ_A then implies that for $\gamma_A' > \gamma_A''$ sufficiently close to γ_A'',

$$\Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A') = \int_{\gamma_A''}^{\gamma_A'} \frac{\partial \Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A)}{\partial \gamma_A} d\gamma_A < 0,$$

while when $\gamma_A' < \gamma_A''$ sufficiently close to γ_A'',

$$\Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A') = -\int_{\gamma_A'}^{\gamma_A''} \frac{\partial \Delta U_P(\beta_P', \gamma_P', \beta_A', \gamma_A)}{\partial \gamma_A} d\gamma_A < 0,$$

which proves the result when $|R_P - R_A| > 0$ is sufficiently small.

For $|R_P - R_A| \to 1$, we have

$$\begin{aligned}
\lim_{|R_P - R_A|\to 1} U_P(q_P(P), q_A(P), P) &= \tilde{q}_P(P)\tilde{u}_P(\pi_P, s_P) - \frac{\phi_P}{2}\left(\tilde{q}_P(P)\right)^2 \\
&= \max_q \left\{ q\tilde{u}_P(\pi_P, s_P) - \frac{\phi_P}{2} q^2 \right\} \\
&> \max_q \left\{ (1 - \tilde{q}_A(A)) q\tilde{u}_P(\pi_P, s_P) - \frac{\phi_P}{2} q^2 \right\} \\
&= (1 - \tilde{q}_A(A))\tilde{q}_P(A)\tilde{u}_P(\pi_P, s_P) - \frac{\phi_P}{2}\left(\tilde{q}_P(A)\right)^2 \\
&= \lim_{|R_P - R_A|\to 1} U_P(q_P(A), q_A(A), A),
\end{aligned}$$

where the variables and functions with a tilde are the limits of their respective variables and functions as $|R_P - R_A| \to 1$. The first and last equalities follow from the product rule for limits using the fact that $\tilde{u}_P(\pi_A, s_A) = 0$. The second and third equalities follow from the fact that the limits of the first-order conditions that define $\tilde{q}_P(d)$ solve these optimization problems. While the inequality follows from the fact that $\tilde{q}_A(A) > 0$.

Continuity of the principal's expected utility in the model parameters, taking as given the delegation decision, then proves the result for the case when the wedge in relative pro-social preferences is large. □

Proof of Proposition 4. Observe that ΔU_P has the same sign as

$$\hat{\Delta} U_P = \Delta U_P \left(2\sqrt{\beta_A^2 + \gamma_A^2}(\beta_A\beta_P - \phi_A\phi_P)^2 \right).$$

Furthermore, when $\gamma_P = 0$, then

$$\frac{\partial \hat{\Delta} U_P}{\partial \gamma_A} = \frac{\beta_P^2 \gamma_A \left(2\beta_A^2(\beta_P - \phi_P) + 4\phi_A\phi_P\sqrt{\beta_A^2 + \gamma_A^2} - 4\beta_A\phi_A\phi_P - 3\gamma_A^2\phi_P \right)}{\sqrt{\beta_A^2 + \gamma_A^2}},$$

which is zero for at most two $\gamma_A > 0$. To show this, replace γ_A^2 by $C^2 - \beta_A^2$ and notice that the term in the brackets in the numerator yields a quadratic equation in C.

At $\gamma_A = 0$, we get that

$$\hat{\Delta} U_P = 0 \quad \textit{and} \quad \frac{\partial \hat{\Delta} U_P}{\partial \gamma_A} = 0 \quad \textit{and} \quad \frac{\partial^2 \hat{\Delta} U_P}{\partial \gamma_A^2} = 2\beta_A\beta_P^2(\beta_P - \phi_P) < 0.$$

Therefore, in a neighbourhood above $\gamma_A = 0$ we have that $\frac{\partial \hat{\Delta} U_P}{\partial \gamma_A} < 0$ and $\hat{\Delta} U_P < 0$.

Furthermore, for $\gamma_A \to \sqrt{\phi_A^2 - \beta_A^2}$, we have that

$$\lim_{\gamma_A \to \sqrt{\phi_A^2 - \beta_A^2}} \hat{\Delta} U_P = -\beta_P^2(\beta_A - \phi_A)(2\beta_A^2\beta_P - 3\beta_A\phi_A\phi_P + \phi_A^2\phi_P),$$

which is positive if $8\beta_P > 9\phi_P$, in which case $\hat{\phi}_A = 0$, or if

$$\phi_A > \hat{\phi}_A = \frac{1}{2}\beta_A \left(3 + \sqrt{9 - \frac{8\beta_P}{\phi_P}}\right).$$

We thus have that: *i*) as $\gamma_A \to 0$, $\hat{\Delta} U_P$ is negative and decreasing in γ_A, *ii*) $\frac{\partial \hat{\Delta} U_P}{\partial \gamma_A} = 0$ has at most two solutions for $\gamma_A > 0$, and *iii*) if $\phi_A > \hat{\phi}_A$, then $\lim_{\gamma_A \to \sqrt{\phi_A^2 - \beta_A^2}} \hat{\Delta} U_P > 0$. Therefore, $\hat{\Delta} U_P$ crosses zero once and we define $\hat{\gamma}_A \in \left(0, \sqrt{\phi_A^2 - \beta_A^2}\right)$ as this point of crossing.

For $\gamma_A \in (0, \hat{\gamma}_A)$, the principal delegates the control rights because $\hat{\Delta} U_P < 0$ and therefore $\Delta U_P < 0$. For $\gamma_A > \hat{\gamma}_A$, the principal retains the control rights because $\hat{\Delta} U_P > 0$ and therefore $\Delta U_P > 0$. □

Proof of Proposition 5. First, from Proposition 4 it follows that the agent follows a threshold delegation strategy where below $\hat{\gamma}_A$, the principal delegates the control rights while above it the principal retains the control rights. From Proposition 1 it then follows that, in each of these two regions, the organization's sustainability is weakly increasing. Furthermore, from the proof of Lemma 2 it follows that $q_A(P) < q_A(A)$ at $\hat{\gamma}_A > 0$ because $R_P = 0 < R_A$. Therefore, the organization's sustainability jumps downwards at $\hat{\gamma}_A$. □

Proof of Proposition 6. Recall that we assume that the delegation set is convex, that is, $\{\gamma_A | d = A\} = [\underline{\gamma}_A, \overline{\gamma}_A]$, where $0 \leq \underline{\gamma}_A \leq \overline{\gamma}_A$. Denote $\gamma_A^* = \{\gamma_A | R_P = R_A\}$. From Proposition 2 and the fact that if indifferent, the principal delegates the control rights to the agent, it follows that $\gamma_A^* \in [\underline{\gamma}_A, \overline{\gamma}_A]$. We first show that $\gamma_A^* < \overline{\gamma}_A$. Assume that $\gamma_A^* = \overline{\gamma}_A$. We know that there exists an $\epsilon > 0$ such that any $\gamma_A \in (\overline{\gamma}_A, \overline{\gamma}_A + \epsilon)$ is a feasible parameter value in that $\gamma_A \geq 0$ and $\phi_A > \max_{(\pi,s)\in\mathcal{P}} u_A(\pi, s)$. Therefore, $\gamma_A^* = \overline{\gamma}_A$ contradicts Proposition 3, which shows that in a neighborhood around γ_A^*, the principal delegates the control rights. As consequence, we must have $\gamma_A^* < \overline{\gamma}_A$.

In the following, we distinguish between two cases. First, assume that $\gamma_A^* = \underline{\gamma}_A$. From the previous argument we know that $\underline{\gamma}_A = \gamma_A^* < \overline{\gamma}_A$. If $\gamma_A^* = \underline{\gamma}_A > 0$, then any $\gamma_A \in [0, \underline{\gamma}_A)$ is feasible. Therefore, $\gamma_A^* = \underline{\gamma}_A > 0$ contradicts Proposition 3, which shows that in a neighborhood around γ_A^*, the principal delegates the control rights. As a result, when $\gamma_A^* = \underline{\gamma}_A$, then $\underline{\gamma}_A = 0$, and the only shift in the control rights takes place at $\overline{\gamma}_A$. At $\overline{\gamma}_A$, $R_A > R_P$ and the control rights are withdrawn from the agent. According to the proof of Lemma 2, this shift in the control rights leads to a reduction in the agent's effort, $q_A(P) < q_A(A)$, and increase in the principal's effort, $q_P(P) > q_P(A)$. These effort changes result in a drop in the agent's effective control, $e_A(P) < e_A(A)$, and therefore in a drop in the organization's sustainability.

Second, assume that $\gamma_A^* \in (\underline{\gamma}_A, \overline{\gamma}_A)$. The same arguments as in the first case show that at $\overline{\gamma}_A$, the shift in the control rights leads to a decrease in the organization's sustainability. Similar arguments show that if $\underline{\gamma}_A > 0$, then at $\underline{\gamma}_A$, the shift in the control rights leads to a decrease in the organization's sustainability. □

Proof of Proposition 7. Observe ΔU_P has the same sign as

$$\hat{\Delta} U_P = \Delta U_P \frac{2(\beta_A \beta_P - \phi_A \phi_P)^2}{\beta_A}.$$

Furthermore, when $\gamma_A = 0$, then

$$\frac{\partial \hat{\Delta} U_P}{\partial \gamma_P} = 2\gamma_P \left(\frac{\beta_A \beta_P^2}{\sqrt{\beta_P^2 + \gamma_P^2}} - \beta_A \phi_P - \frac{2\beta_P \phi_A \phi_P}{\sqrt{\beta_P^2 + \gamma_P^2}} + 2\phi_A \phi_P \right),$$

which is zero for at most one $\gamma_P > 0$ because the function in brackets is monotonic in γ_P.

At $\gamma_P = 0$, we get that

$$\hat{\Delta} U_P = 0 \quad \textit{and} \quad \frac{\partial \hat{\Delta} U_P}{\partial \gamma_P} = 0 \quad \textit{and} \quad \frac{\partial^2 \hat{\Delta} U_P}{\partial \gamma_P^2} = 2\beta_A(\beta_P - \phi_P) < 0.$$

Therefore, in a neighbourhood above $\gamma_P = 0$, we have that $\frac{\partial \hat{\Delta} U_P}{\partial \gamma_P} < 0$ and $\hat{\Delta} U_P < 0$.

Furthermore, as $\gamma_P \to \sqrt{\phi_P^2 - \beta_P^2}$, we have that

$$\lim_{\gamma_P \to \sqrt{\phi_P^2 - \beta_P^2}} \hat{\Delta} U_P = -(\beta_P - \phi_P)^2(-2\phi_A \phi_P + \beta_A(2\beta_P + \phi_P)),$$

which is positive if $\phi_P > \hat{\phi}_P = -((2\beta_A \beta_P)/(\beta_A - 2\phi_A))$.

We thus have that: *i*) as $\gamma_P \to 0$, $\hat{\Delta} U_P$ is negative and decreasing in γ_P, *ii*) $\frac{\partial \hat{\Delta} U_P}{\partial \gamma_P} = 0$ has at most one solution for $\gamma_P > 0$, and *iii*) if $\phi_P > \hat{\phi}_P$, then $\lim_{\gamma_P \to \sqrt{\phi_P^2 - \beta_P^2}} \hat{\Delta} U_P > 0$. Therefore, $\hat{\Delta} U_P$ crosses zero once and we define $\hat{\gamma}_P \in \left(0, \sqrt{\phi_P^2 - \beta_P^2}\right)$ as this point of crossing.

For $\gamma_P \in (0, \hat{\gamma}_P)$, the principal delegates the control rights because $\hat{\Delta} U_P < 0$ and therefore $\Delta U_P < 0$. For $\gamma_P > \hat{\gamma}_P$, the principal retains the control rights because $\hat{\Delta} U_P > 0$ and therefore $\Delta U_P > 0$. □

Proof of Proposition 8. First, from Proposition 7 it follows that the agent follows a threshold delegation strategy where below $\hat{\gamma}_P$, the principal delegates the control rights, while above it she retains the control rights. From the symmetry in the model and Proposition 13 it follows that $e_P(A)$ is increasing in γ_P. Furthermore, from the symmetry in the model and Proposition 14 it follows that when $d = P$ and $\gamma_A = 0$, then $\tilde{\gamma}_P = 0$ and therefore $e_P(P)$ is increasing in γ_P. As a result, in each of these two regions, the organization's sustainability is increasing. Furthermore, from the proof of Lemma 2 it follows that $q_P(A) < q_P(P)$ at $\hat{\gamma}_P > 0$ because $R_A = 0 < R_P$. Therefore, the organization's sustainability jumps upwards at $\hat{\gamma}_P$. □

Proof of Proposition 9. The proof works analogous to the proof of Proposition 6. □

III Proofs for Section III

Proof of Corollary 1. We need to find an $\alpha_s \geq -\gamma_A$ such that

$$R_A^s = \frac{(\gamma_A + \alpha_s)^2}{\beta_A^2 + (\gamma_A + \alpha_s)^2} = R_P, \tag{A.3}$$

where $R_P \in [0, 1)$. Observe that R_A^s is strictly increasing in α_s, for $\alpha_s = -\gamma_A$ we have that $R_A^s = 0$, and $\lim_{\alpha_s \to \infty} R_A^s = 1$. Therefore, there exists a unique α_s such that equation (A.3) is satisfied, meaning that the agent's effective relative pro-social preferences are the same as the principal's relative pro-social preferences and therefore the stakeholders' preferred projects are the same. □

Proof of Proposition 10. From Corollary 1, we know that there exists an $\tilde{\alpha}_s \in [-\gamma_A, \infty)$ such that $R_A^s = R_P$. Given that changing α_s in the extended model is equivalent to changing the agent's pro-social preferences γ_A in the baseline model, it follows from Proposition 3 that for α_s slightly below $\tilde{\alpha}_s$, the principal delegates the control rights.

Furthermore, for $\gamma_A = 0$ (i.e., $\alpha_s = -\gamma_A$ in the extended model), we have

$$\lim_{\beta_P \to 0} \Delta U_P = -\frac{\beta_A \gamma_P^2 (\beta_A - 2\phi_A)}{2\phi_A^2 \phi_P} > 0,$$

and therefore the principal retains the control rights when α_s and β_P are sufficiently small.

Therefore, there exists an $\hat{\alpha}_s < \tilde{\alpha}_s$ where for $\hat{\alpha}_s$, $R_A^s < R_P$, such that delegation switches from the principal to the agent as α_s increases. At this threshold, as discussed in the proof of Lemma 2, the organization's sustainability drops because of the strictly increased effort by the agent. □

Proof of Proposition 11. From Proposition 4 it follows that for $\gamma_A \in (0, \hat{\gamma}_A)$, the principal delegates the control rights while for $\gamma_A \in \left(\hat{\gamma}_A, \sqrt{\phi_A^2 - \beta_A^2}\right)$ the principal retains the control rights.

When $\gamma_P = 0$, then

$$\frac{\partial U_P(q_P(P), q_A(P), P)}{\partial \gamma_A} = 0.$$

Furthermore, when $\gamma_P = 0$, then

$$\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} = \frac{\beta_P^2 \gamma_A \left(-\beta_A^3 \beta_P + \gamma_A^2 \phi_P \left(\sqrt{\beta_A^2 + \gamma_A^2} - \phi_A\right) + \beta_A^2 \phi_P \sqrt{\beta_A^2 + \gamma_A^2}\right)}{\left(\beta_A^2 + \gamma_A^2\right)^{3/2} (\beta_A \beta_P - \phi_A \phi_P)^2}.$$

For $\gamma_A > 0$, this function has the same sign as

$$g(\gamma_A) = -\beta_A^3 \beta_P + \gamma_A^2 \phi_P \left(\sqrt{\beta_A^2 + \gamma_A^2} - \phi_A\right) + \beta_A^2 \phi_P \sqrt{\beta_A^2 + \gamma_A^2}.$$

Substituting $\gamma_A = \sqrt{C^2 - \beta_A^2}$, we get

$$\tilde{g}(C) = -\beta_A^3 \beta_P + C^2(C - \phi_A)\phi_P + \beta_A^2 \phi_A \phi_P.$$

This function has at most three C for which it is zero (because it is a third-order polynomial). The same is then true for $g(\gamma_A)$.

Furthermore, $g(0) > 0$ and $g(\phi_A) > 0$, and therefore $g(\gamma_A)$ has at most two solutions such that $g(\gamma_A) = 0$ for $\gamma_A \in \left[0, \sqrt{\phi_A^2 - \beta_A^2}\right]$. A third is not possible because it contradicts the fact that $g(0) > 0$ and $g(\phi_A) > 0$. A single solution that crosses zero is not possible since $g(0) > 0$ and $g(\phi_A) > 0$. No solution or a single solution that reaches zero but does not cross it is not possible because this would imply that $\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} > 0$ for (almost all) $\gamma_A \in \left(0, \sqrt{\phi_A^2 - \beta_A^2}\right)$, which contradicts the fact that for $\gamma_A > \hat{\gamma}_A$, we have $\Delta U_P > 0$ and the principal retains the control rights (see Proposition 4). Therefore, $g(\gamma_A) = 0$ has two solutions in the interval $\left[0, \sqrt{\phi_A^2 - \beta_A^2}\right]$, which we denote by $\tilde{\gamma}_A$ and $\gamma'_A > \tilde{\gamma}_A$. As a consequence, $\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A}$ is strictly positive for $\gamma_A \in (0, \tilde{\gamma}_A)$ and $\gamma_A \in \left(\gamma'_A, \sqrt{\phi_A^2 - \beta_A^2}\right)$, and strictly negative for $\gamma_A \in (\tilde{\gamma}_A, \gamma'_A)$.

Observe, that the threshold above which the principal delegates the control rights, $\hat{\gamma}_A$, must satisfy $\hat{\gamma}_A \in (\tilde{\gamma}_A, \gamma'_A)$. It cannot be the case that $\hat{\gamma}_A \leq \tilde{\gamma}_A$. The reason is that for $\gamma_A \in (0, \tilde{\gamma}_A]$, we have $\Delta U_P < 0$ because $\frac{\partial U_P(q_P(P), q_A(P), P)}{\partial \gamma_A} = 0$ and $\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} > 0$. Furthermore, it cannot be the case that $\hat{\gamma}_A \geq \gamma'_A$ because for $\gamma_A = \sqrt{\phi_A^2 - \beta_A^2}$, we have $\Delta U_P > 0$ (see Proposition 4) and for $\gamma_A \in \left(\gamma'_A, \sqrt{\phi_A^2 - \beta_A^2}\right]$, we have $\frac{\partial U_P(q_P(P), q_A(P), P)}{\partial \gamma_A} = 0$ while $\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} > 0$, which implies that $\Delta U_P > 0$ for $\gamma_A \in \left(\gamma'_A, \sqrt{\phi_A^2 - \beta_A^2}\right]$. Therefore, $\hat{\gamma}_A \in (\tilde{\gamma}_A, \gamma'_A)$.

For $\gamma_A \in (0, \tilde{\gamma}_A)$, the principal delegates the control rights because $\hat{\gamma}_A \in (\tilde{\gamma}_A, \gamma'_A)$ and therefore $\frac{\partial \max_{d \in \{P,A\}} U_P(q_P(d), q_A(d), d)}{\partial \gamma_A} = \frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} > 0$. For $\gamma_A \in [\tilde{\gamma}_A, \hat{\gamma}_A)$, the principal still delegates the control rights but $\frac{\partial \max_{d \in \{P,A\}} U_P(q_P(d), q_A(d), d)}{\partial \gamma_A} = \frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} \leq 0$. At $\hat{\gamma}_A$, the principal is indifferent between delegating or retaining the control rights and $\frac{\partial U_P(q_P(A), q_A(A), A)}{\partial \gamma_A} \leq 0$ and $\frac{\partial U_P(q_P(P), q_A(P), P)}{\partial \gamma_A} = 0$. For $\gamma_A > \hat{\gamma}_A$, the principal retains the control rights and therefore $\frac{\partial \max_{d \in \{P,A\}} U_P(q_P(d), q_A(d), d)}{\partial \gamma_A} = \frac{\partial U_P(q_P(P), q_A(P), P)}{\partial \gamma_A} = 0$, which proves the first result of the proposition.

The result that the organization's sustainability improves for $\gamma_A \in (0, \tilde{\gamma}_A)$ follows directly from Proposition 1 and the fact that $\tilde{\gamma}_A \leq \hat{\gamma}_A$. □