Sustainable Organizations*

Thomas Geelen† Jakub Hajda‡ Jan Starmans§

December 9, 2022

Abstract

We develop a theory of sustainability-driven stakeholder governance to study how stakeholders with pro-social preferences influence organizations' sustainability—their social outcomes. While stronger pro-social preferences render a stakeholder's project choice more sustainable and increase her incentives to control it, other stakeholders respond by limiting her control rights when conflicts of interest become too severe. This response can lead to an overall decline in organizations' sustainability even when stakeholders become more pro-social.

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

JEL Classifications: D23, G30, L20, Q56.

*We would like to thank Martin Boyer, Alvin Chen, Deeksha Gupta, Martin Hellwig, Marcin Kacperczyk, Doron Levit, Georgii Riabov, Bogdan Stacescu (discussant), Paul Voss, Paul Whelan, and seminar participants at Copenhagen Business School, the University of Bonn, Johns Hopkins University, and the 2022 Young Scholars Nordic Finance Workshop 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.

Organizations face increasing demands from stakeholders such as investors, customers, employees, and regulators to address environmental, social, and governance (ESG) concerns (see, e.g., Dimson et al., 2015; Bolton and Kacperczyk, 2021; Dimson et al., 2021). These demands have been shaping organizations in a significant way. For example, firms representing more than a third of the global market capitalization have science-based targets to reduce carbon emissions (SBTi, 2021), roughly 90% of the S&P 500 and 70% of the Russel 1000 companies publish ESG reports (McKinsey, 2022), and an increasing fraction of shareholder proposals in recent years concerned environmental and social issues (Bolton et al., 2020; Bubb and Catan, 2022). These recent developments have invigorated the debate on the stakeholder society (see, e.g., Bebchuk and Tallarita, 2020; Hart and Zingales, 2022) and raise the question of how and to what extent stakeholders can influence organizations' sustainability efforts and outcomes.

To address this question, we develop a theory of sustainability-driven stakeholder governance. We consider an organization with a monetary and a social payoff and endow its stakeholders with pro-social preferences. In modelling control over the organization's project choice, we follow Aghion and Tirole (1997) and distinguish between control rights and effective control.[1] Even without control rights, the non-controlling stakeholder (e.g., an employee) can have effective control over the organization's project choice by exerting effort and becoming better informed than the controlling stakeholder (e.g., a firm owner). In addition, the controlling stakeholder can delegate control rights to the non-controlling stakeholder. We use this framework to study the economic forces that determine the extent to which socially responsible stakeholders influence the organization. Our goal is to understand whether more pro-social stakeholders are able to exert greater influence on the organization and how conflicts of interests between different stakeholders—driven by differences in pro-social preferences—affect the organization's sustainability efforts and outcomes.

In our framework, the pro-social preferences of both stakeholders affect the choice of the project undertaken by the organization and therefore its ability to generate a social payoff. A key force in our model is that strengthening the pro-social preferences of a stakeholder not only tilts that stakeholder's project choice towards social outcomes, but also affects all stakeholders' incentives to acquire information about project payoffs and therefore their influence

[1] While Aghion and Tirole (1997) call control rights and effective control formal authority and real authority, respectively, we follow the terminology of Burkart et al. (1997).

on the project choice—their *effective control*. We show that strengthening the pro-social preferences of a stakeholder leads to a substitution of effective control between the stakeholders. In particular, effective control can shift to the less pro-social stakeholder. Moreover, an increase in the pro-social preferences of the non-controlling stakeholder can result in a loss of control rights, which can reduce the organization's expected social payoff—the *organization's sustainability*. In addition, we explore the implications of our results for social compensation and agent selection. Overall, we provide a positive theory of stakeholder capitalism that sheds light on how sustainability-driven stakeholders affect organizations. Our theory has implications for the interactions between, among others, managers and employees, boards and managers, as well as entrepreneurs and investors.

In our model, an organization is composed of two risk-neutral stakeholders: a principal and an agent. The principal is the controlling stakeholder (e.g., a manager or firm owner) and the agent is a non-controlling stakeholder (e.g., an employee or investor). There exists a set of projects, one of which can be implemented by the organization. Each potential project has a monetary and a social payoff. Stakeholders may have pro-social preferences in addition to monetary incentives, meaning that they may care about the organization's social payoff. Project payoffs are ex ante unknown. Each stakeholder decides how much effort to exert to learn about the unknown projects' payoffs. The stakeholder with control rights ultimately determines the project choice. However, if this stakeholder is uninformed, then she grants effective control to the other stakeholder if he is informed. At the beginning of the game, the principal decides whether to delegate control rights over the project choice to the agent or to retain them.

We solve the model by backward induction, starting with the project choice. A stakeholder's project choice amounts to allocating the organization's resources to the production of a monetary payoff and a social payoff. Allocating more resources towards the production of the monetary payoff implies a lower social payoff and vice versa. The allocation, and therefore the project choice, is determined by a stakeholder's *relative pro-social preferences*—the strength of pro-social preferences relative to the size of monetary incentives. Stronger relative pro-social preferences lead to a larger allocation of resources towards the production of the social payoff. In particular, we show that pro-social preferences are necessary to generate a conflict of interest between stakeholders in our model, as in the absence of pro-social preferences both stakeholders prefer to allocate all resources to the production of the monetary payoff.

As the next step, we study the principal's and agent's effort choices taking the delegation decision as given. In our model, these effort choices are strategic substitutes. The more effort the principal exerts, the lower the agent's incentives to exert effort and vice versa. Furthermore, we show that changes in a stakeholder's pro-social preferences or monetary incentives lead to a substitution of effort between the two stakeholders and therefore to a substitution of effective control.

We then examine how changes in pro-social preferences affect effective control. We demonstrate that strengthening a stakeholder's pro-social preferences can either increase or decrease the stakeholder's effective control. To see this, note that making a stakeholder more pro-social can increase the incentives to exert effort because stronger pro-social preferences imply a higher utility level. However, it can also reduce the conflict of interest between the two stakeholders, which weakens the incentives to exert effort because losing effective control to the other stakeholder becomes less costly. The overall effect on effort therefore depends on the relative strength of these two forces.

As the final step, we study the delegation of control rights. Holding control rights strengthens a stakeholder's incentives to exert effort because it increases the likelihood of obtaining effective control. Stronger incentives to exert effort give the principal a reason to delegate control rights to the agent even if there is a conflict of interest between the two. Specifically, by doing so, the principal reduces her effort and therefore her effort cost. In addition, delegating control rights to the agent may reduce the risk of no project being implemented. However, it comes at a cost, as it is now more likely that the agent's preferred project will be implemented. We show that when the wedge in relative pro-social preferences between the stakeholders is small, then the benefits of delegation outweigh the costs and the principal delegates control rights to the agent. In particular, the principal may even delegate control rights to a more pro-social agent. In contrast, when this wedge is large, the principal retains control rights.

The delegation of control rights also plays a crucial role when considering the organization's sustainability. The reason is that pro-social preferences not only affect the stakeholders' effort and project choices but also the delegation decision. The general insight is that even a small change in a stakeholder's pro-social preferences can alter the delegation decision and thus cause a large change in the organization's sustainability. The reason is that control rights are delegated to or withdrawn from the more pro-social stakeholder. For example, taking the delegation decision as given, we demonstrate that increasing the agent's pro-social preferences unambiguously raises his effective control when the principal has no pro-social preferences. In addition, it makes his project choice more pro-social. As a result, it improves the organization's sustainability. However, when the conflict of interest between the stakeholders becomes too severe, the principal withdraws control rights from the agent, which reduces the organization's sustainability. This result shows that an increase in a stakeholder's pro-social preferences can hurt the organization's sustainability.[2]

We further extend our model to examine the role of social compensation—compensation based on the organization's social payoff—and agent selection (e.g., hiring, investor selection, internal promotions) on incentives and outcomes. We show that while social compensation can in principle reduce the conflict of interest between the stakeholders and increase their incentives to allocate resources towards the production of the social payoff, it may not necessarily benefit the organization's overall sustainability. Social compensation alters the conflict of interest between the stakeholders, which may cause the principal to change his delegation decision and therefore reduce the organization's sustainability. Finally, we show that if the principal can choose between agents that ex ante differ in their pro-social preferences, then she prefers a more pro-social but not too pro-social agent. The reason is that a more pro-social agent has stronger incentives to exert effort, which benefits the principal. However, as the agent becomes too pro-social, he allocates too many resources to the social payoff, which hurts the principal. In addition, we show that the selection of a more but not too pro-social agent improves the organization's sustainability.

Model Implications. Our model applies to, and has empirical implications for, 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 results provide a theoretical underpinning for the increasing prevalence of shareholder activism and engagement—shareholders trying to obtain effective control—in relation to

[2]In our baseline model, the allocation of control rights is a discrete choice. While control rights are discrete in many real-life applications, our result does not rely on the fact that the change in control rights is discrete. As we illustrate in Appendix C, what we need is that strengthening the agent's pro-social preferences leads to a significant reduction in the agent's control rights.

sustainability issues (see, e.g., Eccles and Klimenko, 2019).

Investors and Managers: One of the broader insights of our model that making stakeholders more pro-social can increase or decrease an organization's sustainability is in line with the mixed empirical evidence regarding the impact of stakeholders on the sustainability footprint of firms and other organizations. For example, Heath et al. (2021) document that socially responsible investment funds do not improve the behavior of their portfolio companies in terms of environmental and social outcomes, despite the intention to generate positive impact. Similarly, Kim et al. (2022) show that ESG-linked loans are not associated with real and positive improvements in corporate ESG practices and that they can sometimes even lead to a deterioration in ESG scores. In contrast, several other studies document a positive effect of stakeholders on firms' sustainability (Di Giuli and Kostovetsky, 2014; Chen et al., 2020; Huang et al., 2021).

Boards and CEOs: Our framework can also be applied to the relationship between a CEO and the board. We show that when the conflict of interest between them increases, the board may eventually stop delegating control rights to the CEO, which can be broadly interpreted as a dismissal. This insight 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 disagreement in terms of pro-social preferences may also induce forced turnover. For example, in 2021, Danone's CEO Emmanuel Faber was removed from his position after his attempt to transform Danone into a company that not only focuses on profits but also on environmental sustainability.[3]

Managers and Employees: A general implication of our analysis is that paying for social performance may be misguided if social payoffs of an organization reflect the choices of multiple stakeholders, as is the case in our model. Intuitively, if the social payoff of an organization 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 designing incentive schemes for managers or employees based on ESG key performance indicators (KPIs), which have become more prevalent in recent years (see, e.g., Ikram et al., 2019; Semler Brossy, 2021; Cohen et al., 2022; Rajan et al., 2022).

Related Literature. Our paper provides a theory of sustainability-driven stakeholder

[3]See "A Top CEO Was Ousted After Making His Company More Environmentally Conscious. Now He's Speaking Out," Time Magazine, 21 November 2021.

society (see, e.g., Tirole, 2001; Allen et al., 2015; Magill et al., 2015) and contributes to several strands of the literature. First, our analysis relates to the literature in organizational economics that studies how control rights affect their holders' incentives to acquire costly information to exercise those rights (see, e.g., Aghion and Tirole, 1997; Stein, 2002).[4] We contribute to this literature by considering organizations that generate both monetary and social payoffs and by endowing stakeholders with pro-social preferences. Introducing project choice over two dimensions—monetary and social payoffs—and endowing agents with preferences over both dimensions—monetary incentives and pro-social preferences—is necessary to study how pro-social preferences affect control in organizations and how control in turn translates into the organization's social performance.

Second, our paper adds to the theoretical literature on corporate governance.[5] While the existing literature focuses on shareholders and corporate executives, our paper applies to a broader set of stakeholders which are important in corporate governance such as employees, mid-level managers, customers, non-equity investors, and regulators. Furthermore, there exists little formal analysis of the impact of pro-social stakeholders on corporate governance, with the work by Matsusaka and Shu (2021), Gollier and Pouget (2022), and Levit et al. (2022) on shareholder voting being notable exceptions. Studying pro-social stakeholders is, however, important in light of the growing empirical evidence highlighting the crucial role of sustainability concerns in the context of corporate governance (see, 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).

Finally, we add to the growing theoretical literature that studies the impact of socially responsible stakeholders on organizations, which has thus far focused on investors and firms (see, 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] We provide a theory of sustainability-

[4] A related literature studies how control rights affect the communication between agents in organizations (see, 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 (see, e.g., Burkart et al., 1997; Van den Steen, 2010a,b; Chen, 2022).

[5] See Malenko (2022) for a survey.

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

driven stakeholder engagement that applies to a broader set of stakeholders, beyond investors, and a broader set of organizations, beyond firms, such as universities and NGOs. Understanding the interaction between stakeholders regarding the organization's sustainability is important given the increasing demands from stakeholders to address social and environmental concerns.[7]

I Model

We consider an organization composed of a risk-neutral principal P (she) and a risk-neutral agent A (he). We refer to the principal and the agent jointly as the organization's stakeholders and to the principal as the controlling stakeholder. For example, the principal is a manager of a firm and the agent is an employee or the principal is an entrepreneur and the agent is an investor. There is a set of projects that differ in their social and monetary payoffs, one of which can be implemented by the organization. There are three dates without time discounting. At time zero, the principal decides whether to delegate control rights over the project choice to the agent. At time one, both the principal and the agent decide on their effort to learn about the set of available projects. At least one stakeholder needs to be informed for a project to be implemented. At time two, a project may be implemented. We describe the model in more detail below.

The output of the organization is described by 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 are two production technologies. The first production technology generates highly negative monetary and social payoffs irrespective of $\iota \in [0, 1]$. The second production technology generates a monetary payoff of $\pi = \sqrt{1-\iota}$ and a social payoff of $s = \sqrt{\iota}$. As such, the relevant set of projects that a stakeholder may be willing to implement

[7]See, for example, "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.

is generated by the second production technology and is given by

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

There is no savings technology such that the organization generates a zero monetary payoff and a zero social payoff if the initial resources of the organization are not employed.

The two stakeholders face an informational friction because they cannot distinguish between the two production technologies without acquiring additional information. Furthermore, they believe that the production technology that generates highly negative payoffs makes up a positive fraction of all projects. As a result, if neither the principal nor the agent is informed, then no project is implemented. Both stakeholders can exert costly effort to learn about the production technologies. Specifically, the principal chooses a probability $q_P \in [0,1]$—the principal's effort—with which she learns about the two production technologies. Her private cost of effort is $\frac{\phi_P}{2} q_P^2$. If the principal is informed, then she can distinguish between the two production technologies and can choose a project (π, s) from the set $\mathcal{P}$. Similarly, the agent chooses a probability $q_A \in [0,1]$—the agent's effort—with which he learns about the two production technologies and his private cost of effort is $\frac{\phi_A}{2} q_A^2$. The stakeholders make their effort choices simultaneously and the outcomes of the principal's effort and the agent's effort are independent.

The principal's utility from a project (π, s) is

$$u_P(\pi, s) = \beta_P \pi + \gamma_P s,$$

where $\gamma_P \geq 0$ captures the principal's pro-social preferences and $\beta_P > 0$ captures her monetary incentives. Similarly, the agent's utility from a project (π, s) is

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

where $\gamma_A \geq 0$ captures the agent's pro-social preferences and $\beta_A > 0$ captures his monetary incentives. It is well documented that some stakeholders have pro-social preferences and that they differ in these preferences (e.g., Hong and Kacperczyk, 2009; Hong and Kostovetsky, 2012; Gibson et al., 2021).[8] To ensure an interior equilibrium in the stakeholders' effort

[8]There exists ample evidence in the economics literature that some economic agents have pro-social

choices, we assume that effort is sufficiently costly: $\phi_P > \max_{(\pi,s)\in\mathcal{P}} u_P(\pi, s)$ and $\phi_A > \max_{(\pi,s)\in\mathcal{P}} u_A(\pi, s)$.

Finally, at time zero, the principal decides on the delegation of control rights d over the organization's project choice. The principal either retains *control rights*, $d = P$, or delegates them to the agent, $d = A$. The stakeholder holding control rights decides on the organization's project choice at date two but can also delegate the organization's project choice to the other stakeholder ex post. We refer to the stakeholder whose preferred project is implemented as the stakeholder holding *effective control*.

Figure 1 summarizes the timeline of the model. First, the principal decides whether to delegate 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. There are two frictions in our model: the informational friction regarding the project's payoffs and the non-contractibility of stakeholders' actions.

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' preference regarding short-term and long-term profits. It should be noted that 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.

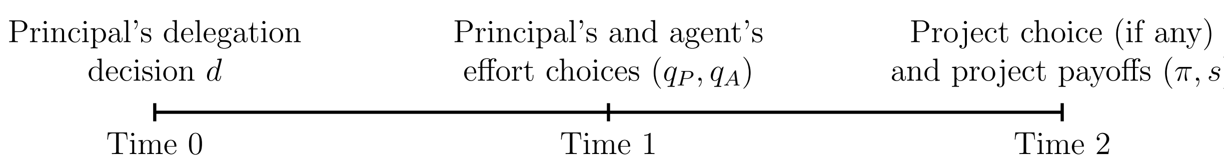


Figure 1: **Model Timeline.**

II Equilibrium

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

preferences and that there is heterogeneity across agents in this dimension (see, e.g., List et al., 2009) the principal's delegation decision.

A Project Choice

At time two, either the principal or the agent holds control rights. If the principal holds control rights and is informed, then she chooses the investment in the social payoff ι such that

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

which yields $\iota_P = \frac{\gamma_P^2}{\gamma_P^2+\beta_P^2}$. Thus, the principal's preferred project is given by $(\pi_P, s_P) = \left(\sqrt{1-\iota_P}, \sqrt{\iota_P}\right)$. We refer to $R_P = \frac{\gamma_P^2}{\gamma_P^2+\beta_P^2} \in [0,1)$ as the principal's *relative pro-social preferences*. Similarly, if the agent has control rights and is informed, then he chooses ι such that $\max_{\iota\in[0,1]} u_A\left(\sqrt{1-\iota}, \sqrt{\iota}\right)$, which yields $\iota_A = \frac{\gamma_A^2}{\gamma_A^2+\beta_A^2}$. Thus, the agent's preferred project is given by $(\pi_A, s_A) = \left(\sqrt{1-\iota_A}, \sqrt{\iota_A}\right)$. We refer to $R_A = \frac{\gamma_A^2}{\gamma_A^2+\beta_A^2} \in [0,1)$ as the agent's relative pro-social preferences. Note that each stakeholder's utility from implementing the other stakeholder's preferred project is strictly positive because $\beta_P > 0$ and $\beta_A > 0$.[9]

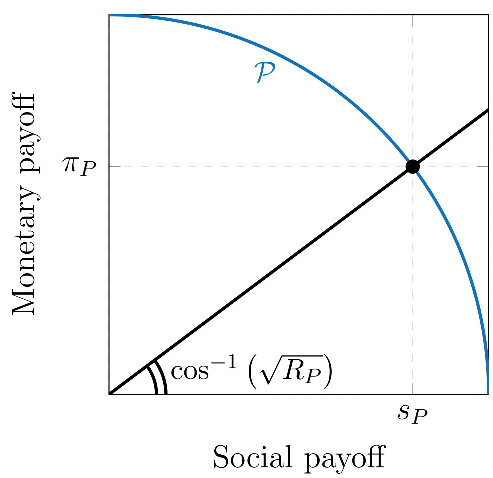
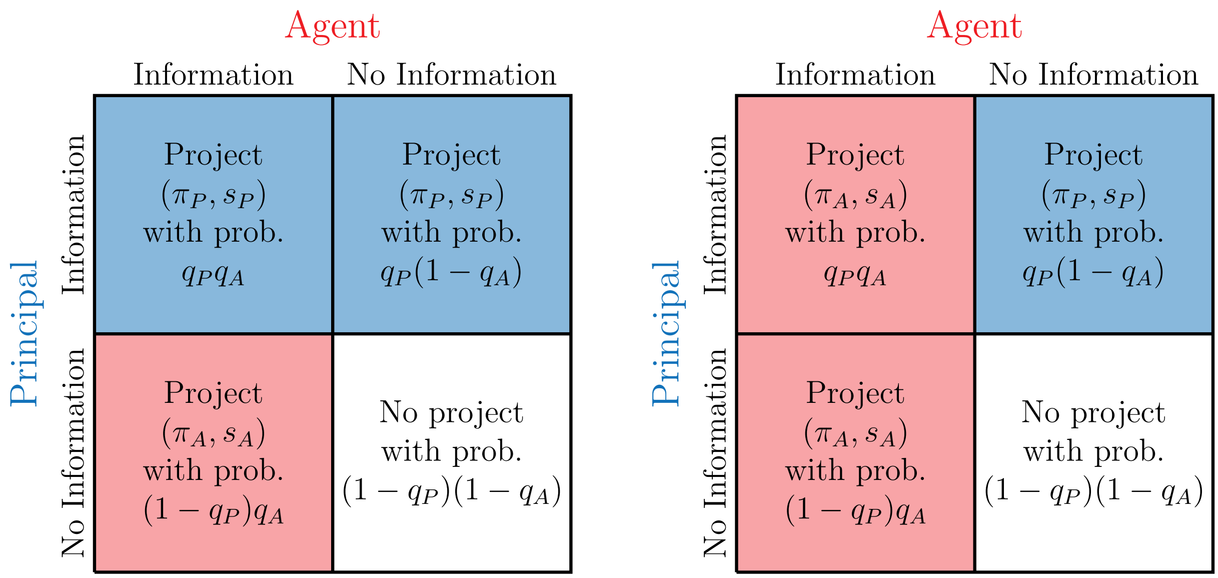


Figure 2: **Set of Projects and Principal's Preferred Project.** The figure plots the set of projects $\mathcal{P}$ and the principal's preferred project (π_P, s_P) as a function of the principal's relative pro-social preferences $R_P = \frac{\gamma_P^2}{\gamma_P^2+\beta_P^2}$.

Figure 3 summarizes the allocation of effective control and the project that is implemented as a function of the allocation of control rights. If the principal has control rights and is

[9]The stakeholders' relative pro-social preferences satisfy $R_P < 1$ and $R_A < 1$ because $\beta_P > 0$ and $\beta_A > 0$. As a result, we have $\pi_P > 0$ and $\pi_A > 0$ while $s_P \geq 0$ and $s_A \geq 0$. In particular, we have $u_P(\pi_A, s_A) \geq \beta_P \pi_A > 0$ and $u_A(\pi_P, s_P) \geq \beta_A \pi_P > 0$.



(a) Principal has control rights $d = P$.

(b) Agent has control rights $d = A$.

Figure 3: **Control Rights and Effective Control.**

informed, then she implements her preferred project (π_P, s_P). If the principal has control rights and is not informed but the agent is, then she follows the agent's recommendation—the agent has effective control—and implements the agent's preferred project (π_A, s_A) because $u_P(\pi_A, s_A) > 0$. If neither stakeholder is informed, then no project is implemented due to the risk of generating highly negative payoffs. If the agent has control rights and is informed, then his preferred project (π_A, s_A) is implemented, while the principal's preferred project (π_P, s_P) is implemented if the agent is not informed but the principal is.

We assume that the principal cannot renege on her 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 that is not the case, this 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 control rights. For example, Baker et al. (1999) micro-found this reputational cost in a repeated delegation-of-authority model.

B Effort

Given the project choice at time two, we can determine the stakeholders' expected utilities at time one, which in turn determine their effort choices. The principal's expected utility is given by

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

If the principal holds control rights, $d = P$, then she has effective control with probability q_P. The agent has effective control with probability $(1-q_P)q_A$, which happens when he is informed while the principal is not. Delegating control rights to the agent, $d = A$, means that the principal has effective control with a lower probability $(1-q_A)q_P$ while the agent has it with a higher probability q_A.

Similarly, the agent's expected utility is given by

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

The two stakeholders choose their effort levels simultaneously. Thus, the equilibrium effort choices (conditional on the delegation decision d) are determined by the two first-order conditions

$$\frac{\partial U_P(q_P, q_A, d)}{\partial q_P} = (1-q_A) u_P(\pi_P, s_P) + \mathbb{I}_{\{d=P\}} q_A \Delta u_P - \phi_P q_P = 0,$$
$$\frac{\partial U_A(q_P, q_A, d)}{\partial q_A} = (1-q_P) u_A(\pi_A, s_A) + \mathbb{I}_{\{d=A\}} q_P \Delta u_A - \phi_A q_A = 0,$$

where $\Delta u_P = u_P(\pi_P, s_P) - u_P(\pi_A, s_A) \geq 0$ and $\Delta u_A = u_A(\pi_A, s_A) - u_A(\pi_P, s_P) \geq 0$.

The principal's and agent's best-response functions are thus given by

$$B_P(q_A, d) = \frac{(1-q_A) u_P(\pi_P, s_P) + \mathbb{I}_{\{d=P\}} q_A \Delta u_P}{\phi_P},$$
$$B_A(q_P, d) = \frac{(1-q_P) u_A(\pi_A, s_A) + \mathbb{I}_{\{d=A\}} q_P \Delta u_A}{\phi_A}.$$

In particular, we have $\frac{\partial B_P(q_A,d)}{\partial q_A} < 0$ and $\frac{\partial B_A(q_P,d)}{\partial q_P} < 0$. That is, the principal's and agent's effort choices are strategic substitutes. Intuitively, if the stakeholder holding control rights exerts more effort, then the likelihood that the other stakeholder gains effective control decreases and so do the incentives to exert effort. However, if the stakeholder without control rights exerts more effort, then the other stakeholder is less concerned about being uninformed. This happens because the option to transfer effective control when being uninformed works as a hedge for the stakeholder who holds control rights. Increasing the value of this option weakens the incentives to exert effort.

In addition, holding control rights increases the incentives to exert effort, which results from two effects. First, the likelihood of having effective control is higher, which induces more effort. Second, the option to transfer control rights when uninformed reduces the incentives to exert effort. However, because a stakeholder obtains a higher utility while having effective control, the first effect dominates the second.

Lemma 1 (Equilibrium in 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 utility and the agent's utility with respect to their effort levels.*

C Delegation of Control Rights

Finally, the principal decides whether to delegate control rights, taking into account her own as well as the agent's future actions, 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 from Lemma 1.

III Model Analysis

Having characterized the solution to our model, we next study how the principal's and agent's monetary incentives and pro-social preferences affect project choice, effort, as well as control rights and effective control. In addition, we investigate how changes in pro-social preferences influence the expected social payoff of the organization.

A Project Choice

Monetary incentives and pro-social preferences affect project choice. The principal's investment in the organization's social payoff is determined by her relative pro-social preferences, that is, $\iota_P = R_P$. Similarly, the agent's investment in the organization's social payoff is determined by his relative pro-social preferences, that is, $\iota_A = R_A$.

Corollary 1 (Pro-social Preferences and Project Choice)**.** *Stronger relative pro-social preferences lead to lower monetary payoffs and higher social payoffs. That is, for $j \in \{P, A\}$,*[10]

$$\frac{\partial \pi_j}{\partial R_j} < 0 \quad \textit{and} \quad \frac{\partial s_j}{\partial R_j} > 0.$$

A stakeholder with stronger relative pro-social preferences is willing to accept a lower monetary payoff in order to generate a higher social payoff and therefore invests more in the social payoff and less in the monetary payoff. For example, an employee with stronger pro-social preferences is 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).

Corollary 2 (No Pro-social Preferences and Project Choice)**.** *Without pro-social preferences, $\gamma_P = \gamma_A = 0$, the principal's and agent's preferred projects coincide, that is, $\iota_P = \iota_A = 0$ and $(\pi_P, s_P) = (\pi_A, s_A) = (1, 0)$.*

Both stakeholders prefer to allocate all resources to the production of the monetary payoff because $\gamma_P = \gamma_A = 0$. Corollary 2 implies that pro-social preferences are necessary to drive a wedge between the stakeholders' preferred projects. In particular, even if $\beta_P \neq \beta_A$,

[10] In this corollary, we use the extended real line that includes $\pm\infty$ to account for the fact that $\frac{\partial s_j}{\partial R_j} = +\infty$ if $R_j = 0$.

their preferred projects coincide in the absence of pro-social preferences. Thus, differences in monetary incentives alone are insufficient to generate a conflict of interest between the stakeholders in our model.

Corollary 3 (Wedge in Relative Pro-social Preferences and Project Choice)**.** *Increasing the wedge between the stakeholders' relative pro-social preferences, $|R_P - R_A|$, increases the wedge between their investments in the organization's monetary and social payoffs, $|\iota_P - \iota_A|$.*

Corollary 3 shows that pro-social preferences can create a conflict of interest between the principal and the agent if the two stakeholders don't share the same relative pro-social preferences. In particular, whenever $R_P \neq R_A$, the two stakeholders prefer different projects and therefore different levels of the organization's social payoff.

Lemma 1 implies that each stakeholder can determine the organization's project choice with a positive probability. Put differently, each stakeholder has effective control with some probability. As a result, both the principal's and the agent's relative pro-social preferences affect the organization's expected social payoff.

Corollary 4 (Project Implemented by Organization)**.** *When the principal and the agent have different relative pro-social preferences, $|R_P - R_A| > 0$, then the organization sometimes undertakes the principal's preferred project and sometimes the agent's, which differs from the principal's preferred project.*

Because both stakeholders in our model determine the organization's choices and payoffs rather than only the controlling stakeholder, our framework constitutes a novel theory of the sustainability-driven stakeholder society. We provide a positive theory of the stakeholder society in which conflicts of interest between stakeholders arise due to differences in pro-social preferences and in which all stakeholders affect the organization's choices and payoffs. The question we address in the following subsections is what determines the extent to which the different stakeholders influence the monetary and social payoffs of the organization.

B Effort, Effective Control, and the Organization's Sustainability

Having discussed project choice, we next study the determinants of stakeholders' effort choices. In addition, we analyze the allocation of effective control resulting from stakeholders' effort choices. Finally, we study the organization's sustainability—the expected social payoff. The analysis in this subsection takes the delegation decision as given, which we study in the next subsection.

As discussed in Subsection II.B, the principal's and agent's effort choices are strategic substitutes. That is, an increase in the agent's effort reduces the principal's incentives to exert effort and vice versa. However, changes in the principal's or agent's monetary incentives or pro-social preferences may affect both the principal's and the agent's incentives to exert effort. As the following result shows, changes in the stakeholders' incentive and preference parameters still lead to a substitution between the principal's and the agent's effort levels in equilibrium.

Proposition 1 (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* $\theta \in \{\beta_P, \gamma_P, \beta_A, \gamma_A\}$*,*

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

We next translate the effort levels into the allocation of effective control, as it partly determines the organization's monetary and social payoffs. 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, and denote it by $e_P(d)$.[11] The agent's allocation of effective control is $e_A(d) = 1 - e_P(d)$. The substitution of effort between the principal and agent from Proposition 1 implies a substitution of effective control between the principal and agent in line with the changes in effort.

Corollary 5 (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* $\theta \in \{\beta_P, \gamma_P, \beta_A, \gamma_A\}$*,*

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

[11]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 $e_P(A) = \frac{(1-q_A(A))q_P(A)}{q_P(A)-q_P(A)q_A(A)+q_A(A)}$.

Proposition 1 and Corollary 5 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 influences the project choice.

Recent evidence suggests that some stakeholders are increasingly acknowledging the social outcomes of their actions and thus becoming more pro-social (see, e.g., Morgan Stanley, 2019). 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.[12]

Proposition 2 (Pro-social Preferences and Effort when the Principal Holds Control Rights)**.** *When the principal holds control rights, $d = P$, then she reduces her effort and the agent increases his effort when the agent's pro-social preferences γ_A increase, that is,*

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

Strengthening the agent's pro-social preferences unambiguously increases his incentives to exert effort. Figure 4 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 more about the organization's social payoff, his utility when he has effective control increases, which in turn increases his incentives to exert effort.

Importantly, the principal's best response function and therefore her incentives to exert

[12]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. Higher effort cost ϕ_P of the principal decreases her incentives to exert effort and therefore lowers her 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.

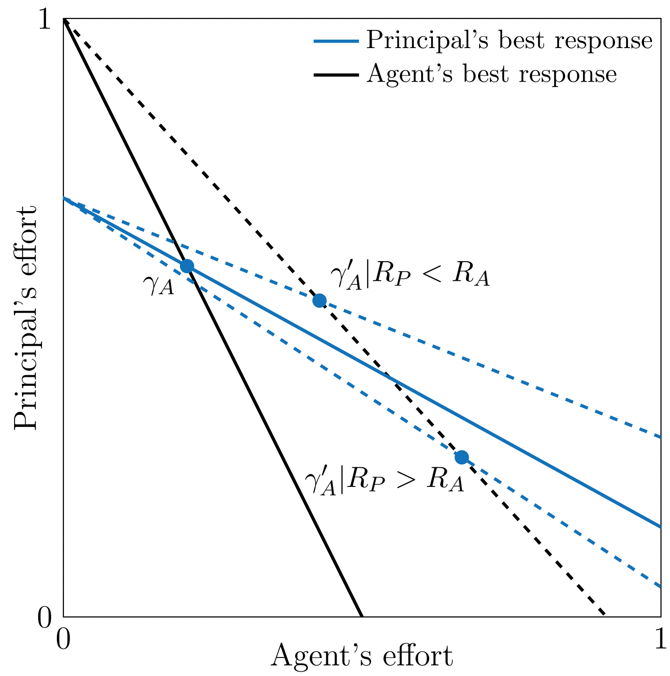


Figure 4: **Agent's Pro-social Preferences and Equilibrium Effort when the Principal Holds Control Rights.** The figure plots the principal's best response function $B_P(q_A, P)$ and the agent's best response function $B_A(q_P, P)$. The solid lines are the best response functions for some initial level 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 $R_P < R_A$.

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 project choice 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 4. 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 results in them gaining more effective control and therefore more influence on the firm's outcomes.

Proposition 3 (Pro-social Preferences and Effort when the Agent Holds Control Rights)**.** *When the agent holds control rights, $d = A$, then there exists a threshold $\bar{\gamma}_A$ such that when $\gamma_A < \bar{\gamma}_A$, 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 $\gamma_A > \bar{\gamma}_A$, 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 3 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 control rights. As is clear from the principal's best response function, $B_P(q_A, A)$, her 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 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 his incentives to exert effort, depends on his 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 he fails 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 $\gamma_A = \bar{\gamma}_A$, the direct utility effect starts to dominate.[13] A further increase in γ_A thus leads to the agent exerting more effort and to the principal exerting less effort.

The general insight of the results regarding the allocation of effective control is that increasing stakeholders' pro-social preferences can lead to more influence of these stakeholders on organizations. As such, our results provide a theoretical underpinning for the increasing prevalence of shareholder activism in relation to sustainability issues (see, e.g., Eccles and Klimenko, 2019).

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 $d = P$ ($d = A$).

Given that changes in the stakeholders' pro-social preferences alter their project choices and the allocation of effective control, it is crucial to understand how those changes jointly affect the organization's social outcomes. We analyze this by looking at the expected social

[13] The condition $(1 - q_P(A))s_A + q_P(A)(s_A - s_P) = 0$ implicitly characterizes $\bar{\gamma}_A$ 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 $\bar{\gamma}_A$ must be below the level at which $R_P = R_A$. Thus, the threshold $\bar{\gamma}_A$ satisfies $R_A < R_P$.

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 sustainability comes at the cost of a lower monetary payoff.

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

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

There are two effects at play. First, making the agent more pro-social shifts effective control. This effect is positive when effective control shifts to the most pro-social stakeholder. As shown in Propositions 2 and 3, the agent's effective control increases unless the agent holds control rights and $\gamma_A < \bar{\gamma}_A$ in which case it decreases. The second effect is that making the agent more pro-social tilts his project choice towards a higher social payoff, which has a positive effect on the organization's sustainability. Note that due to the symmetry of the model, similar trade-offs arise when we strengthen the principal's pro-social preferences.

Proposition 4 (Pro-social Preferences and Organization's Sustainability)**.** *When the principal has no pro-social preferences, $\gamma_P = 0$, then the organization's sustainability, conditional on the delegation decision, is increasing in the agent's pro-social preferences*

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

Proposition 4 demonstrates that when the principal has no pro-social preferences, and as the agent becomes more pro-social, then, conditional on the delegation decision, 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 4 shows that more pro-social stakeholders can make an organization more sustainable, this result takes the delegation decision as given. However, this decision ultimately belongs to the principal and it may change as stakeholders become more pro-social. We study this last crucial step in the following subsection.

C Delegation of Control Rights

The delegation decision plays a vital role in determining the organization's outcomes. We first study its impact on stakeholders' effort choices. As discussed in Subsection II.B, giving a stakeholder control rights increases the stakeholder's incentives to exert effort. As the following proposition shows, this increases the stakeholder's equilibrium effort level.

Lemma 2 (Control Rights and Effort)**.** *Delegating 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* $q_A(A) \geq q_A(P)$*.*

Importantly, the delegation decision does not affect the stakeholders' effort choices if they have no pro-social preferences.

Corollary 6 (No Pro-social Preferences, Control Rights, and Effort)**.** *Without pro-social preferences,* $\gamma_P = \gamma_A = 0$*, the principal's and agent's effort choices are independent of the delegation decision, that is,* $(q_P(P), q_A(P)) = (q_P(A), q_A(A))$*.*

More generally, even if stakeholders have pro-social preferences, the principal's and agent's project choices are the same when there is no conflict of interest, that is, when $R_P = R_A$ (Corollary 3). The first-order conditions for the stakeholders' effort choices then imply that the incentives to exert effort are independent of the delegation decision, that is,

$$(q_P(P), q_A(P)) = (q_P(A), q_A(A)) \quad \Leftrightarrow \quad R_P = R_A.$$

Pro-social preferences, therefore, serve two roles in our model. First, they are necessary to drive a wedge between the stakeholders' project choices (Corollary 2). Second, they are necessary for the delegation decision to affect stakeholders' effort choices. Put differently, the delegation decision influences effective control only if stakeholders have different relative pro-social preferences.

To study the principal's delegation decision, we define the wedge in the principal's utility between retaining and delegating formal 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 control rights (i.e., $d = A$) if $\Delta U_P < 0$ and retains control rights (i.e., $d = P$) if $\Delta U_P > 0$.

Proposition 5 (Irrelevance of Control Rights)**.** *When the stakeholders have the same relative pro-social preferences, $R_P = R_A$, then the delegation decision d is irrelevant for the effort choices and payoffs. In particular, $\Delta U_P = 0$.*

Proposition 5 shows that if there is no conflict of interest between the stakeholders, then the principal's expected utility does not depend on the delegation decision. The result therefore highlights that pro-social preferences and the wedge in stakeholders' relative pro-social preferences are necessary to make the delegation decision relevant. A direct consequence of this result is that the delegation decision is irrelevant when neither stakeholder has pro-social preferences.

Corollary 7 (No Pro-social Preferences and Irrelevance of Control Rights)**.** *Without pro-social preferences, $\gamma_P = \gamma_A = 0$, the delegation decision d is irrelevant for effort and payoffs. In particular, $\Delta U_P = 0$.*

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

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

We express the utility as the probability that a project is implemented times the expected utility conditional on a project being implemented, minus the effort cost.

Equation (1) shows that the delegation decision affects the expected utility in three ways. The first effect of delegating 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 control rights. This, in turn, is beneficial to the principal. Intuitively, while the principal knows that her expected utility is lower when the agent is in control, delegating control rights to the agent may increase the probability that a project gets implemented. In other words, delegating control rights to the agent may reduce the risk of no project being implemented.

The second effect of delegating control rights to the agent is an *effort cost effect*. The principal reduces her effort and the agent increases his effort, which lowers the principal's effort cost by

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

always giving the principal an incentive to delegate control rights.

The third effect of delegating control rights to the agent is that it alters the types of projects undertaken. As doing so increases the agent's effort and reduces the principal's effort, it also strengthens the agent's effective control. As a consequence, we have that

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

Intuitively, transferring effective control to the agent means that his 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 control rights to the agent.

Proposition 6 (Relative Pro-social Preferences Wedge 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* *control rights to the agent, $d = A$. If the wedge is relatively large, then the principal retains control rights, $d = P$.*

Intuitively, if the conflict of interest between the principal and the agent is small, 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 utility loss for the principal. In this case, the value the principal obtains from the agent's increased effort and her lower effort cost dominates and therefore the principal delegates control rights. In contrast, if the conflict of interest is significant, then the project selection effect becomes larger and the principal retains 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 get the following result.[14]

Proposition 7 (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 $\hat{\gamma}_A$ such that the principal delegates control rights when $\gamma_A \in (0, \hat{\gamma}_A)$ and she retains control rights when $\gamma_A > \hat{\gamma}_A$.*

Recall from Proposition 4 that the organization's sustainability conditional on the delegation decision is increasing in γ_A when $\gamma_P = 0$. The crucial insight from Proposition 7 is that increasing the agent's pro-social preferences can lead to a withdrawal of control rights from the agent. The following result shows that this withdrawal can actually have negative consequences for the organization's sustainability.

Proposition 8 (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 7, that is,*

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

[14]The condition $\phi_A > \hat{\phi}_A$ in Proposition 7 ensures that if γ_A gets sufficiently large, then the principal wants to retain formal control rights.

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

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

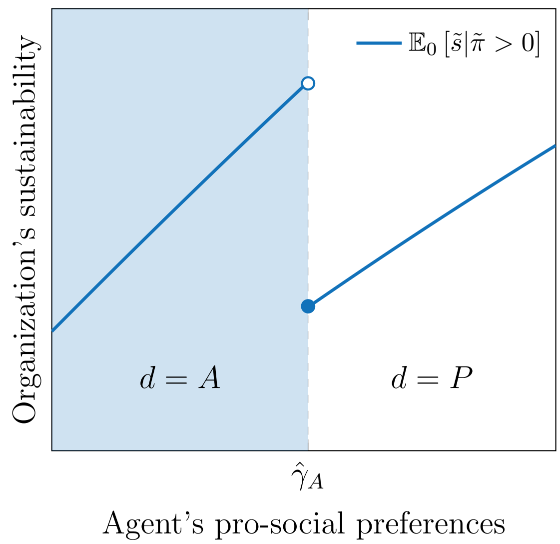


Figure 5: **Agent's Pro-social Preferences and Organization's Sustainability.** The figure plots the organization's sustainability, $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$, as a function of the agent's pro-social preferences γ_A. If $\gamma_A < \hat{\gamma}_A$, the principal delegates control rights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

Figure 5 shows the overall effect of increasing the agent's pro-social preferences from Proposition 8, taking into account the endogenous delegation decision by the principal. The plot highlights the discrete reduction in the organization's sustainability as the principal withdraws control rights. For example, if investors become more socially responsible, they may lose control rights and therefore have less influence on the firm, which can ultimately hurt the firm's sustainability.

In many real-life applications, control rights are discrete. For example, control of a firm changes discreetly around the majority voting threshold. However, our result in Proposition 8 that an increase in the agent's pro-social preferences can reduce the organization's sustainability due to a withdrawal of control rights does not rely on the fact that the change in control rights is discrete. As we illustrate formally in Appendix C, what we need is that strengthening the agent's pro-social preferences leads to a significant reduction in the agent's control rights.

The insight that the principal may withdraw control rights from the agent as the conflict of interest between them aggravates is in line with the findings of Huang et al. (2020), who show that disagreement between investors and management is an important driver of CEO turnover, which can be broadly interpreted as the withdrawal of control rights. Anecdotal evidence suggests that disagreement in terms of pro-social preferences may also induce turnover. For example, in 2021, Danone's CEO Emmanuel Faber was removed from his position after the attempt to transform Danone into a company that not only focused on profits but also on environmental sustainability.[15]

The broader implication of our results in this section is that making stakeholders in an organization more pro-social may not necessarily increase the organization's sustainability. This is in line with the mixed evidence in the literature regarding the impact of stakeholders on the sustainability footprint of firms and other organizations. For example, Heath et al. (2021) document that socially responsible investment funds do not improve the behavior of their portfolio companies in terms of environmental and social outcomes, despite the intention to generate positive impact. Similarly, Kim et al. (2022) show that ESG-linked loans are not associated with real and positive improvements in corporate ESG practices and are associated with a deterioration in ESG scores in some cases. In contrast, several other studies document a positive effect of stakeholders on firms' sustainability (see, e.g., Di Giuli and Kostovetsky, 2014; Chen et al., 2020; Gibson et al., 2021; Huang et al., 2021; Noh and Oh, 2021).

We now turn our attention to the principal becoming more pro-social. We obtain a result similar to Proposition 7 regarding the delegation decision when the agent has no pro-social preferences.

Proposition 9 (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 $\hat{\gamma}_P$ such that the principal delegates control rights when $\gamma_P \in (0, \hat{\gamma}_P)$ and she retains control rights when $\gamma_P > \hat{\gamma}_P$.*

If the conflict of interest between the stakeholders is small, then the principal delegates control rights to the agent because this increases the agent's effort and reduces the probability of no project being implemented. If the conflict of interest is significant, then the cost of

[15] See "A Top CEO Was Ousted After Making His Company More Environmentally Conscious. Now He's Speaking Out," Time Magazine, 21 November 2021.

delegating control rights is too high and the principal retains control rights.[16]

In this case, the organization's sustainability is always increasing in the principal's pro-social preferences. The reason is that at $\hat{\gamma}_P$, the control rights get taken away from the less pro-social agent, which improves the organization's sustainability.

Proposition 10 (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 continuously for all $\gamma_P \neq \hat{\gamma}_P$, where $\hat{\phi}_P$ and $\hat{\gamma}_P$ are the thresholds from Proposition 9, 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,

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

Overall, the results in this subsection highlight that small changes in pro-social preferences can lead to large changes in organization's sustainability.

IV Extensions

In this section, we study several extensions of our model. Specifically, we study the role of social compensation and agent selection (e.g., hiring, investor selection, internal promotions).

A Social Compensation

In this subsection, we introduce 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 subsection is that while social compensation can reduce the conflict of interest between stakeholders, it can also hurt the organization's sustainability through shifts in control rights.

[16]For example, in 2022, HSBC Asset Management's head of responsible investing was suspended after giving a controversial speech entitled "Why Investors Need Not Worry About Climate Risk." He left the bank shortly after. See "HSCB Banker Quits Over Climate Change Furore," Financial Times, 7 July 2022.

To begin with, the analysis in Subsection II.B implies that paying for social performance may be misguided. The reason is that an organization's sustainability reflects not only the principal's choices, but also the agent's (Corollary 4). 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 to the design of incentive schemes for managers or employees based on ESG KPIs, which have become more prevalent: currently as many as 57% of the S&P 500 firms evaluate their managers' performance based on ESG metrics (see, e.g., Ikram et al., 2019; Semler Brossy, 2021; Cohen et al., 2022; Rajan et al., 2022).

We extend the baseline model from Section I by giving the agent 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 $\iota_A = R_A^s$. The first simple implication of introducing social compensation is that if the principal can flexibly adjust the agent's social compensation, then it allows her to eliminate any conflict of interest.

Corollary 8 (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 are 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 not reduce them. In addition, when α_s is positive, then compensating the agent is costly for the organization. As such, 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. Moreover, social payoffs may 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. We can show that, in the context of our model, once there are multiple social payoffs and heterogeneity in preferences across stakeholders regarding the different dimensions of social payoffs, then social compensation based on a single rating that measures the overall social performance of the organization is often insufficient to eliminate the conflict of interest between stakeholders.

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 are discussed in Subsection III.B. We can also use the results from Propositions 6 and 7 to study the effects of social compensation on the delegation decision and the organization's sustainability. When the wedge in effective relative pro-social preferences $|R_P - R_A^s|$ is small, then the principal delegates control rights to the agent, while when the wedge is large, then she retains control rights. This implies that introducing social compensation can align preferences and therefore incentivize the principal to delegate 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 control rights.

Proposition 11 (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 when social compensation crosses this threshold, the organization's sustainability decreases discontinuously, that is,*

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

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

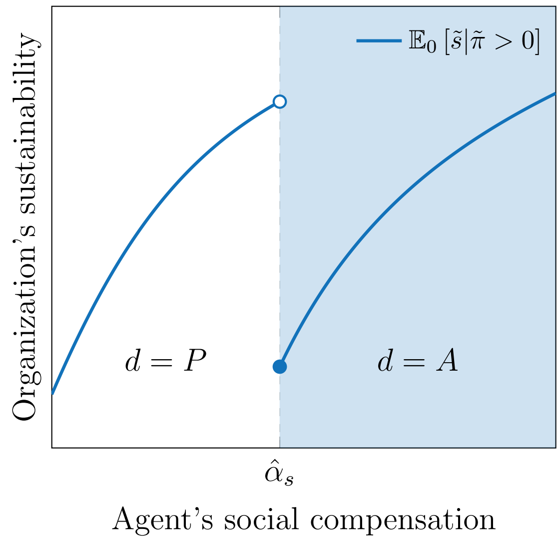


Figure 6: **Effect of Agent's Social Compensation on Organization's Sustainability.** The figure plots the organization's sustainability, $\mathbb{E}_0[\tilde{s}|\tilde{\pi} > 0]$, as a function of the agent's social compensation α_s. As the social compensation grows beyond the threshold $\hat{\alpha}_s$,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

B Agent Selection

In this subsection, we discuss the implications of selecting an agent from a set of agents ex ante (e.g., hiring, investor selection, internal promotions). That is, given that differences in relative pro-social preferences generate a conflict of interest between the principal and the agent, a crucial question to understand is whether the principal has an incentive to select an agent with different relative pro-social preferences to begin with. For example, would a manager without pro-social preferences ever hire a pro-social employee or would an entrepreneur without pro-social preferences raise capital from a socially responsible investor?

One may expect the principal to select an agent with similar pro-social preferences to avoid any conflict of interest. However, as Proposition 12 shows, this may not always be the case.

Proposition 12 (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 if and only if $\gamma_A \in (0, \tilde{\gamma}_A)$, selecting a marginally more pro-social agent improves the principal's expected utility, that is,*[17]

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

Furthermore, for $\gamma_A \in (0, \tilde{\gamma}_A)$, the organization's sustainability also improves, 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 his monetary incentives fixed, the higher the agent's pro-social preferences, the higher his utility u_A, independent of project choice. This, in turn, increases his 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 a too pro-social agent, thereby increasing the organization's sustainability. Our model thus highlights that agent selection can be used a means of improving the sustainability of an organization.

V Conclusion

We develop a theory of sustainability-driven stakeholder governance to study how stakeholders with pro-social preferences influence an organization's sustainability—its social outcomes.

Our framework extends the framework of Aghion and Tirole (1997) by considering an organization which generates both a monetary and a social payoff and by endowing stakeholders with pro-social preferences. In our setup, differences in pro-social preferences drive a wedge between the stakeholders in terms of their preferred projects to be implemented by the organization. We show that strengthening a stakeholder's pro-social preferences may not always result in a more sustainable organization. While stronger pro-social preferences increase the stakeholder's incentives to improve social outcomes, other stakeholders respond

[17] 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 8, where it is only continuous.

by shifting influence and control rights. As such, an organization's sustainability may decline as stakeholders become more pro-social. In particular, we highlight that small changes in pro-social preferences can have a large impact on organizations' sustainability.

We extend our model to analyze how social compensation and agent selection affects organizations. We show that higher social compensation does not always lead to a more sustainable organization. Furthermore, we find that a principal who cares only about the monetary payoff might select a more pro-social agent. This agent selection improves the organization's sustainability. Overall, our theory sheds light on how sustainability-driven stakeholders affect organizations.

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.

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

Bebchuk, L. A. and Tallarita, R. (2020). The illusory promise of stakeholder governance. *Cornell Law Review*, 106:91.

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

Bolton, P. and Kacperczyk, M. T. (2021). Firm commitments. *Working Paper, Columbia Business School.*

Bolton, P., Li, T., Ravina, E., and Rosenthal, H. (2020). Investor ideology. *Journal of Financial Economics*, 137(2):320–352.

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.*

Bubb, R. and Catan, E. M. (2022). The party structure of mutual funds. *Review of Financial Studies*, 35(6):2839–2878.

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.

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.*

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*.

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*.

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*.

Heath, D., Macciocchi, D., Michaely, R., and Ringgenberg, M. C. (2021). Does socially responsible investing change firm behavior? *Working Paper, University of Utah*.

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, 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. et al. (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.

McKinsey (2022). Does ESG really matter—and why?

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

Morgan Stanley (2019). Sustainable signals: Individual investor interest driven by impact, conviction and choice.

Noh, D. and Oh, S. S. (2021). Measuring institutional pressure for greenness: A demand system approach. *Working Paper, Princeton University*.

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., 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*.

SBTi (2021). Science based targets initiative annual progress report 2021.

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.

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.

Appendix

The first part of the appendix derives the results for the baseline model. The second part derives the results for the model extensions. The third part shows robustness of our results.

A Baseline Model

Proof of Lemma 1. 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 formal authority would be better off setting his 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 Corollary 4. From Corollary 3, it follows that the principal and agent undertake different projects, that is, $(\pi_P, s_P) \neq (\pi_A, s_A)$. Lemma 1 shows that $(q_P(d), q_A(d)) \in (0,1)^2$, which proves the result. □

Proof of Proposition 1. 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

$$\beta_A(\beta_P\gamma_A - \beta_A\gamma_P) - \gamma_A\phi_A\phi_P \leq \beta_A(\beta_P\gamma_A - \beta_A\gamma_P) - \gamma_A(\beta_A\beta_P + \gamma_A\gamma_P)$$
$$= -(\beta_A^2 + \gamma_A^2)\gamma_P \leq 0.$$

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 5. 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 2. The result follows directly from the derivations in the proof of Proposition 1. □

Proof of Proposition 3. Given that the agent holds 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 1, $\frac{\partial q_P(A)}{\partial \gamma_A} \leq 0$. Similarly, if $s_A - q_P s_P < 0$, then $\frac{\partial q_A(A)}{\partial \gamma_A} < 0$ and, as implied by Proposition 1, $\frac{\partial q_P(A)}{\partial \gamma_A} \geq 0$. 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 $\bar{\gamma}_A$ such that $s_A - q_P s_P = 0$. For $\gamma'_A > \bar{\gamma}_A$, we have $s_A - q_P s_P > 0$, and for $\gamma'_A < \bar{\gamma}_A$, we have $s_A - q_P s_P < 0$, which proves the result. □

Proof of Proposition 4. When $d = P$, the result follows directly from Proposition 2, Corollary 5, and the fact that

$$\mathbb{E}_0 \left[\tilde{s} | \tilde{\pi} > 0, d\right] = e_A(d) s_A. \tag{A.3}$$

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

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

The result then follows from Proposition 3, Corollary 5, and equation (A.3). □

Proof of Lemma 2. Direct calculations imply that

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

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

Proof of Corollary 6. The result follows directly from the proof of Lemma 2. □

Proof of Proposition 5. From Corollary 3 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 6. 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 5 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

$$\left.\frac{\partial \Delta U_P(\beta'_P, \gamma'_P, \beta'_A, \gamma_A)}{\partial \gamma_A}\right|_{\gamma_A = \gamma''_A} = 0,$$

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

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 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 equality follow from the product rule for limits using the fact that $\tilde{u}_P(\pi_A, s_A) = 0$. The second and third equality 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 when the wedge in relative pro-social preferences is large. □

Proof of Proposition 7. 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

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

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 control rights because $\hat{\Delta} U_P < 0$ and therefore $\Delta U_P < 0$. For $\gamma_A > \hat{\gamma}_A$, the principal retains 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 $\hat{\gamma}_A$, the principal delegates control rights while above it she retains control rights. From Proposition 4 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 9. 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 control rights because $\hat{\Delta} U_P < 0$ and therefore $\Delta U_P < 0$. For $\gamma_P > \hat{\gamma}_P$, the principal retains control rights because $\hat{\Delta} U_P > 0$ and therefore $\Delta U_P > 0$. □

Proof of Proposition 10. First, from Proposition 9 it follows that the agent follows a threshold delegation strategy where below $\hat{\gamma}_P$, the principal delegates control rights while above it she retains control rights. From the symmetry in the model and Proposition 2 it follows that $e_P(A)$ is increasing in γ_P. Furthermore, from the symmetry in the model and Proposition 3 it follows that when $d = P$ and $\gamma_A = 0$, then $\bar{\gamma}_P = 0$ 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 $\hat{\gamma}_P$. □

B Model Extensions

Proof of Corollary 8. We need to find an $\alpha_s \geq -\gamma_A$ such that

$$R_A^s = \frac{(\gamma_A + \alpha_s)^2}{\beta_A^2 + (\gamma_A + \alpha_s)^2} = R_P, \tag{A.4}$$

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.4) 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 11. From Corollary 8, 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 6 that for α_s slightly below $\tilde{\alpha}_s$, the principal delegates 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 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 12. From Proposition 7 it follows that for $\gamma_A \in (0, \hat{\gamma}_A)$, the principal delegates control rights while for $\gamma_A \in \left(\hat{\gamma}_A, \sqrt{\phi_A^2 - \beta_A^2}\right)$ the principal retains 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

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

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 control rights (see Proposition 7). 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 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 $\hat{\gamma}_A \geq \tilde{\gamma}_A$ because for $\gamma_A = \sqrt{\phi_A^2 - \beta_A^2}$, we have $\Delta U_P > 0$ (see Proposition 7) 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 $\frac{\partial U_P(q_P(A),q_A(A),A)}{\partial \gamma_A} > 0$. Therefore, $\hat{\gamma}_A \in (\tilde{\gamma}_A, \gamma'_A)$.

For $\gamma_A \in (0, \tilde{\gamma}_A)$, the principal delegates 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 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 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 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 4 and the fact that $\tilde{\gamma}_A \leq \hat{\gamma}_A$. □

C Robustness

In this section, we show that our result in Proposition 8 that an increase in the agent's pro-social preferences can reduce the organization's sustainability due to a withdrawal of control rights does not rely on the fact that the change in 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 $\frac{1}{N}(\pi_A, s_A)$, 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 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 control rights to the agent when $\Delta U_P^i = \frac{1}{N}(\Delta U_P + \epsilon_i) < 0$ and retains 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 allocates 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 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 8, 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 $\frac{\partial \mathbb{E}_0[\tilde{s}|\tilde{\pi}>0]}{\partial \gamma_A} < 0$.