WORKING PAPERS

N° 1521

March 2024

"Social Responsibility, Consequentialism and Public Policy"

Paul-Henri Moisson

Toulouse School of Economics

Social Responsibility, Consequentialism and Public Policy*

Paul-Henri Moisson†

March 29, 2024

Abstract

What do prosocial behaviours look like when their consequences are determined by market equilibria? How do market forces and government interventions influence the definition of the "morally right" behavior from a consequentialist perspective? The paper investigates socially responsible investment, asking to what extent investors' ethics drive technology investment, environmental outcomes and asset returns. The paper unveils the determinants of a "morality premium" and derives implications for welfare and public policy.

Keywords: Socially responsible investment, consequentialism, impact investing, green premium, Pigou tax, divestment, shareholder activism.

JEL Codes: A13, D62, H23, Q59

---

*The author is indebted to Jean Tirole for invaluable guidance and support. The author is grateful to Sophie Moinas, Sébastien Pouget, Benjamin Roth, François Salanié and to participants at various seminars at TSE for helpful suggestions.

†Toulouse School of Economics (TSE), paulhenri.moisson@tse-fr.eu.

1 Introduction

What are the individual returns and aggregate consequences of prosocial behaviours in a world subject to both market forces and government interventions? Given the latter, what are the prescriptions of different moral criteria? And given the agents' moral compass, what policy interventions can improve welfare? This paper develops a model that tries to capture the feedback loop between moral criteria, market equilibria and government interventions.

For illustrative purposes, it focuses on socially responsible investment (SRI) in the context of environmental degradation.[1] Following a theoretical approach, this paper notably asks whether and when a *green premium* arises; what drives its sign and its magnitude; what taxes can implement the welfare-optimal investments; what the impact of divestment strategies is; when investing in a dirty technology and subsequently mitigating its externalities is the "morally right thing to do".

The answers to these questions hinge on the agents' moral criteria. As a benchmark, this paper considers investors concerned with the actual impact of their portfolio decisions, i.e., investors abiding by *direct consequentialism.* In selecting the consequentialist approach, this paper makes no pretense at descriptive relevance. Indeed, while the appeal to "impact" permeates most narratives in the SRI industry (and in policymaking), evidence suggests that in practice, a large share of investments labelled as "socially responsible" is not predominantly concerned with impact.[2] Rather, this paper investigates what would happen if investors converged on the view that direct impact is what matters for assessing behaviour. After detailing the direct consequentialist case, the paper considers two other moral criteria, comparing their implications with those of direct consequentialism. The working environment is kept as simple as possible to preserve tractability.

The basic framework, developed in Section 2, has two competitive sectors, green and brown.[3] Each sector is characterized by an installed base of plants and by an investment function for new plants. New investment reflects the price that investors are willing to pay for shares. Willingness to pay hinges on two motivations: (i) intrinsic motivation, i.e. the internalization of part of the impact of investment on pollution; (ii) extrinsic motivation, driven

[1] Estimates of the size of "socially responsible" investments widely differ depending on the methodologies. If anything, claims of "social responsibility" are widespread and increasing. In the case of Europe, based on a large sample of 36, 000 EU-domiciled actively-managed investment funds, a study by the European Securities and Markets Authority (2023) finds that 14% of such funds include ESG-related language in their name, representing EUR 974 billion out of EUR 6.8 trillion for the total sample.

[2] Most SRI currently seem to fail the criterion of additionality. For instance, Ehlers et al. (2020) find that green bonds have not necessarily translated into comparatively low or falling carbon emissions at the firm level.

[3] While the brown technology generates a negative environmental externality, the green one generates no such externality.

by the return differential between shares in green and brown firms. Both the impact of one's investment choice and the green premium are equilibrium determined.

The first insight in this baseline model is the existence of a positive green premium whenever investment choices alter brown investment at the margin, i.e. in such a case, there can be no "doing well by doing good" for investors. By contrast, when investment choices do not alter brown investment at the margin, the green premium disappears, in which case green investors do as well as brown investors, but do not do any good. We study how investors' sophistication in assessing the consequences of their investments affect their choices and aggregate outcome. Naive investors over-estimate their impact by ignoring the consequences of their investments on the market prices. Consequently, higher sophistication generates higher pollution and a lower, yet still positive green premium. We provide a simple measure of the wedge between a naive and a sophisticated perception of one's impact.

Policy interventions. We successively consider taxation, (brown) divestment strategies and shareholder activism. We first investigate the optimal taxation of returns on brown investments in Section 3.1. We find that with (direct) consequentialist, sophisticated investors internalizing both the environmental externality and the tax proceeds generated by their saving decision, a utilitarian planner can implement the first-best investments with Pigouvian taxes (with rate equal to the unit externality). By contrast, with naive investors, the optimal tax can be higher or lower than the Pigouvian tax.

While divestment strategies have been advocated as a prominent way of steering prosocial behaviours,[4] their impact on the externality level is still debated. Could a government increase social welfare by mandating public entities to divest brown shares, or would consequentialist investors and frictionless capital markets render such interventions neutral? Section 3.2 studies divestment strategies and characterizes under which scenario they can decrease the pollution level and increase social welfare, and conversely, under which conditions a neutrality result obtains. Even when divestment has an impact on aggregate pollution, it is partially mitigated by the endogeneity of prices: as the financial returns on the brown technology rise following the divestment decision, additional brown investments from non-divestors partially offset the divested ones.[5]

The same technology can in practice often be operated in several ways, yielding different

[4] Divestment strategies arised among the first forms of SRI, notably with the exclusion of "sin stocks". See for instance Landier and Nair (2008).

[5] This market-equilibrium mitigation due to the heterogeneity in the investors' motivation is also a key driver of the trade-offs investors face in the settings of Green and Roth (2023) and Oehmke and Opp (2023).

levels of externalities. For instance, more attention may be devoted to a plant's waste, or some highly polluting processes may be curbed in favour of less polluting ones. Such efforts are nonetheless likely to entail productivity losses, and hence may not be implemented by managers focusing solely on financial returns. Section 3.3 thus asks what the morally right action is when the brown technology can be "cleaned": should highly-moral savers invest in the green technology, or should they rather invest in the brown one and subsequently clean it?[6] We assume that each investor has full decision rights over the production units she owns, and is thus pivotal in implementing change. We show that the moral pecking order between "investing in the green technology" and "investing in the brown technology and cleaning it" depends on the installed capital stocks: from a direct consequentialist perspective, green investments dominate cleaned brown ones if and only if the installed capital stocks are sufficiently low, and are dominated otherwise.[7]

Section 4 asks whether these insights hold under two alternative moral criteria: shared responsibility, by which investors feel responsible for the *average* impact of investments in the same technology, and rule consequentialism, by which investors feel responsible for their induced externality *if everyone mimicked their behavior*. Under *laissez-faire*, shared responsibility always yields lower aggregate pollution than direct consequentialism, while with shared responsibility as with sophisticated direct-consequentialism, the Pigouvian tax maximizes aggregate welfare. The insights obtained under direct consequentialism when new brown investments obtain in equilibrium, e.g. when installed capital stocks are sufficiently low, still hold with shared responsibility and rule consequentialism. However, the *neutrality* results, due to, e.g., large installed capital stocks and by which, e.g., the green premium can be nil and divestment strategies have no impact on the aggregate externality, fail with shared responsibility and rule consequentialism. Hence, while under direct consequentialism, the moral pecking order depends on the setting, it does not under shared responsibility nor under rule consequentialism: investing in the green technology is always the "morally right" action according to these criteria.

Section 5 considers several extensions – image concerns, consumers with ethical concerns, decommissioning of installed capital stocks, and a lower-bound constraint on investors' finan-

[6] Although their settings significantly differ from ours, this question is also at the heart of Gollier and Pouget (2014), Hart and Zingales (2017) and Landier and Lovo (2020), among others.

[7] Indeed, when installed capital stocks are so high that new brown investments are strictly unprofitable for producers, the only way to reduce the aggregate externality is to clean brown stocks. Consequently, investing in brown stocks and cleaning them becomes the *morally right* action from a direct consequentialist perspective.

cial returns. Section 6 discusses the related literature. Section 7 concludes with alleys for future research. Omitted proofs are in the Appendix.

2 Model

2.1 Baseline setting

There are two technologies, "green" (g) and "brown" (b), with installed capital stocks in each technology k_g, $k_b > 0$. Each unit of capital of each technology produces one unit of the consumption good. The production operators are price-takers. For $s \in \{b, g\}$, the profit-maximization of technology-s producers generates a supply of investment i_s in technology s, as a function of the technology-s price, p_s, given by $i_s = h_s(p_s)$. We assume that for $s \in \{b, g\}$, there exists $c_s \geq 0$ such that $h_s(p_s) = 0$ for any $p_s \leq c_s$, and that the supply function h_s is twice continuously differentiable with $h'_s(p_s) > 0$ for $p_s \in [c_s, +\infty)$.[8,9]

Externality. The production of one unit of consumption with the brown technology generates a negative externality e, due to e.g., pollution and environmental harm. The green technology does not generate such an externality. Therefore, the aggregate externality cost is given by: $e(k_b + i_b)$.

Investors' preferences. An investor's willingness-to-do-good v is drawn from a full-support distribution F on $\mathbb{R}_+$, with continuous density f. Investors save a fixed amount (normlaized to one unit of a numéraire) and consume their savings' returns,[10] i.e. $1/p_g$ for green investors and $1/p_b$ for brown ones. Their allocation decision is binary: $a \in \{0, 1\}$ where $a = 1$ represents investing in the green technology. [All our analysis goes through, changing the interpretation of F, if investors save heterogeneous amounts ($F(v)$ being then the total savings of investors with types $v' \leq v$).]

Investors' choice is motivated by two drivers:

(i) *Intrinsic motivation for the common good*, equal to the product of the agent's willingness-to-do-good v and their perception of the (avoided) induced externality, $e^\dagger$ for a brown

[8] In the baseline setting, installed capital stocks cannot be decommissioned. We discuss this possibility in Section 5: while all the insights remain valid, decommissioning adds new regions to the analysis.

[9] The monotonicity of the investment functions stems from the fact that higher prices lower the cost of capital for producers, and hence more investments receive funding. See Section 2.4 for microfoundations of the investment-supply functions h_s.

[10] If one unit of investment has a cost p, then an investor saving 1 buys $1/p$ units of investment, which produce $1/p$ units of the consumption good.

investment, and 0 for a green one.[11]

(ii) *Monetary incentives*, measured by the loss of consumption associated with a green portfolio:

$$-\left(\frac{1}{p_b} - \frac{1}{p_g}\right)$$

We refer to the extra cost of a green portfolio $(1/p_b - 1/p_g)$ as the *"green premium"* and denote it by γ.

Therefore, a type-v investor solves:

$$\max \left\{ e^{\dagger} v - \left(\frac{1}{p_b} - \frac{1}{p_g}\right);\ 0 \right\}$$

As a consequence, in equilibrium, $1/p_b \geq 1/p_g$ and there exists a unique cut-off type v^* such that agents with type $v < v^*$ save in the brown technology, while agents with type $v > v^*$ save in the green technology. Hence, the investors' segmentation is determined by

$$e^{\dagger} v^* - \left(\frac{1}{p_b} - \frac{1}{p_g}\right) = 0. \quad (1)$$

The market clearing conditions of both technologies write:

$$F(v^*) = (k_b + i_b(p_b))p_b \qquad \text{and} \qquad 1 - F(v^*) = (k_g + i_g(p_g))p_g. \quad (2)$$

The equilibrium prices are thus determined by the equilibrium in the equity market (1) and the two market clearing conditions for brown and green portfolios (2). In particular, as F has no atom in 0, we will have that $v^* > 0$ in equilibrium.

Investors' perception of the induced externality $e^{\dagger}$. We assume investors are *direct consequentialists*, i.e. that they are concerned by the actual (environmental) damage generated by their personal investment.[12] In particular, whenever they are not pivotal, i.e. whenever the brown capital unit would have been operated even in the absence of their investment, they do not feel responsible for the induced damage.

For expositional simplicity, we first consider the following *naive* perception of the induced

[11] Investors derive utility from the harm they avoid doing: an investor's motivation thus considers the externality she would induce by switching to the other technology. The shape of $e^{\dagger}$ then depends on the investors' moral criterion (see below).

[12] See Sinnott-Armstrong (2023) for an overview of philosophical issues associated with consequentialism.

externality for an impact investor:[13]

$$e^{\dagger} = \begin{cases} e/p_b & \text{if } p_b \geq c_b, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Investors are thus naive inasmuch as they neglect the (potential) impact of their portfolio choices on the equilibrium prices. We consider in Section 2.3 a more sophisticated assessment of the induced externality, emphasizing that the naive specification (3) overestimates an investor's actual impact.

Notwithstanding, up to an assumption on the brown investment function, our results do not hinge on the investors' degree of sophistication when assessing the impact(s) of their personal investment.

2.2 Equilibrium and comparative statics

The equilibrium is uniquely given by (1)-(2)-(3).

Lemma 1 (Green premium). *The green premium γ is always positive, strictly so if and only if $e^{\dagger} > 0$, i.e. if and only if $[k_b + k_g + h_g(c_b)]c_b \leq 1$. Conversely, $e^{\dagger} = 0$ if and only if $p_g = p_b = p < c_b$ with $[k_b + k_g + h_g(p)]p = 1$.*

By Lemma 1, the green premium is nil whenever the green investments are impactless, and the latter are so whenever there is no additional brown investment on top of the installed capital base. As an illustration, suppose that energy is produced from an (inelastic) installed base of brown technology[14] and from an elastic supply of renewable, clean energy. Then, investing in brown shares has no impact on GHG emissions, and there would be no green premium in a direct-consequentialist world.

Proposition 1 derives the comparative statics of the equilibrium prices p_b, p_g, green premium γ and aggregate pollution $e(k_b + ib)$.

Proposition 1 (Comparative statics for the equilibrium prices, green premium and aggregate pollution).

[13] When $p_b \geq c_b$, were a marginal share of investors to switch from the green to the brown technology, they would induce a strictly positive level of additional brown investment (as $h_b'(p_b) > 0$ for $p_b \in [c_b, +\infty)$). By contrast, whenever $p_b < c_b$, such investors would buy installed brown capital stocks and the total level of brown investments would not be altered. Hence, when $p_b \geq c_b$, a (marginal) green investor switching to the brown technology would buy $1/p_b$ units of brown investment, which would generate a total externality of e/p_b that would not have been produced otherwise; whereas for $p_b < c_b$, the same investor would buy $1/p_b$ units of installed plants, which would have been operated anyway.

[14] With low operating costs and high decommissioning costs – see Section 5 and Appendix D for a discussion of decommissioning.

(i) A higher (unit) externality e induces a lower brown price p_b, a higher green price p_g, and a higher green premium γ, while having an ambiguous impact on aggregate pollution $e(k_b + i_b)$.

(ii) A higher (green or brown) capital stock k_s induces a lower brown price p_b and a lower green price p_g. A higher brown capital stock k_b always induces a higher green premium γ. The impact of a higher green capital stock k_g on the green premium is ambiguous in general. A higher brown capital stock k_b induces higher aggregate pollution, while a higher green capital stock k_g induces lower aggregate pollution.

(iii) A higher (green or brown) investment supply function h_s (considering homothetic variations)[15] induces a lower brown price p_b and a lower green price p_g. A higher brown supply function h_s always induces a higher green premium γ. The impact of a higher green supply function h_g on the green premium is ambiguous in general. A higher brown supply function h_b induces higher aggregate pollution, while a higher green supply function h_g induces lower aggregate pollution.

(iv) A higher distribution F of willingnesses-to-do-good according to first-order stochastic dominance induces a lower brown price p_b, a higher green price p_g, a higher green premium γ and lower aggregate pollution $e(k_b + i_b)$.

A cleaner brown technology (in the sense of a lower unit externality e) does not necessarily yield a lower aggregate level of pollution $e(k_b + i_b)$. Indeed, when the unit externality e decreases, the perceived induced externality $e^\dagger$ decreases, and as a consequence, the level of brown investment may increase. This "morality-rebound" effect parallels the "rebound effect" in energy economics, whereby more energy-efficient technologies may eventually lead to higher consumption.

Let us illustrate some implications of Proposition 1. A higher (unit) externality e – e.g., as new evidence accrues about the harms associated with the brown technology's pollution – or a higher (in first-order stochastic dominance) distribution of willingnesses-to-do-good – e.g., as awareness of the harms associated with the brown technology's pollution increases – generate a more acute moral pressure to save in the green technology, thereby driving the brown price down, the green price up, and yielding a higher green premium.

Following on Proposition 1, to reduce aggregate pollution, a government may contemplate

[15] We consider homothetic variations in the technologies costs: namely, we write the investment functions as $\alpha_g h_g$ and $\alpha_b h_b$ with α_g and α_b two strictly positive constants, and explore the comparative statics with respect to α_g and α_b. Changes in the cost functions may be due for instance to a technological shock or a change in the regulation.

increasing the green capital stock – e.g., by paying for green capital units on its own budget – or decreasing the brown capital stock – e.g., by closing existing brown plants via a stricter environmental regulation. While both interventions indeed reduce aggregate pollution, increasing the green capital stock raises both brown and green prices with an ambiguous impact on the green premium, whereas decreasing the brown capital stock raises both and decreases the green premium. By claim (iii), the same analysis applies to taxing brown supply or subsidizing green supply.

Remark: Naive dynamics. Importantly, the distribution F captures both the distribution of intrinsic motivations, and possibly heterogeneous saving amounts. Whenever the green premium is strictly positive, brown investors get higher financial returns than green ones. Hence, in a future period, brown investors have a higher wealth than green investors, i.e. the distribution of savings and willingnesses-to-do-good shifts to a distribution with a fatter lower tail and a thinner upper tail. Proposition 1 would then suggest that the current green premium pushes the future brown price down, the future green price up and thus the future green premium down. Nonetheless, in a fully dynamic perspective, several other factors may push in other directions – e.g., the evolution of demand may depend on the (dynamic) capital stocks, investors may have intertemporal strategies, etc.

Remark. Models with exogenous material incentives imply that compliance (here, the quantity $1 - F(v^*)$) increases with material incentives (here, the green premium γ). By contrast, this model predicts the opposite correlation: a higher level of compliance (lower v^*) is associated with *lower* material incentives (i.e. a lower green premium) due to the market equilibrium.

2.3 Investors' sophistication

We examine in this section the role of investors' *sophistication* in assessing the consequences of their actions. In particular, we derive an alternative expression for the perceived induced externality under direct consequentialism.

From a direct consequentialist perspective, the perceived induced externality $e^\dagger$ writes differently depending on whether an investor expects other investors with similar types to behave in the same way as she does, which we relate to the investor's degree of sophistication. While in our running specification (3), an investor supposes she would be the only one to change her behaviour, we now investigate the more sophisticated assumption that other investors with similar types change their behaviour too. We show in Appendix A that all the main insights derived with our running specification of the perceived induced externality still hold with this alternative specification, requiring an additional, mild assumption on the brown investment function. In fact, the crucial feature of direct consequentialism is that the perceived induced externality is nil for any $p_b < c_b$, and strictly increases with the number of brown capital units purchased (i.e. strictly decreases with p_b) for $p_b \geq c_b$.

With the sophisticated specification, the induced externality $e^\dagger$ writes as the ratio of the externality generated by a marginal brown investment over the marginal variation in the fraction of brown investors:

$$e^\dagger_{\text{soph}} = \frac{edi_b}{f(v^*)dv^*} \tag{4}$$

Indeed, if types in $[v^*, v^* + dv^*)$ (representing a share $f(v^*)dv^*$ of the population) switch to the brown technology, they generate an externality equal to edi_b. The above equation then gives the *per capita* externality generated by those investors.[16] Hence, we assume that investors care about their impact on the prices, which impact the producers' supply and the number of investment units owned by each investor, but do not take into account their (potential) impact on the technology choices of other investors.

Differentiating the brown market clearing condition yields an explicit expression for the perceived induced externality:

$$e^\dagger_{\text{soph}} = \begin{cases} \dfrac{e}{p_b(1+\eta_b^{-1})} > 0 & \text{if } p_b \geq c_b, \\ 0 & \text{otherwise.} \end{cases}$$

where $\eta_b \equiv p_b h_b'(p_b)/[k_b + h_b(p_b)]$ is the price-elasticity of brown capital.[17]

[16]To keep things simple, we assume that to compute the change in brown investment di_b, investors keep the technology choices of types $[0, v^*)$ and $[v^* + dv^*, +\infty)$ fixed.

[17]Letting $I_b \equiv k_b + i_b$ be the level of brown capital, the induced externality $e^\dagger_{\text{soph}}$ may be rewritten for $p_b \geq c_b$ as

$$e^\dagger_{\text{soph}} = \frac{e}{p_b + \dfrac{dp_b}{dI_b} I_b}.$$

Relatedly, an alternative, equivalent expression for $e^\dagger_{\text{soph}}$ consists in computing the *per capita* impact on total induced externality which is given by $eF(v)/p_b(v)$ where $p_b(v)$ is the equilibrium brown price for cut-off v ($F(v)$ brown investors each owning $1/p_b(v)$ units of brown capital). Therefore,

$$e^\dagger_{\text{soph}} = \frac{1}{f(v^*)dv^*}\left(\frac{F(v^*+dv^*)}{p_b(v^*+dv^*)} - \frac{F(v^*)}{p_b(v^*)}\right).$$

Importantly, this sophisticated perceived induced externality arises as the limit of the induced externality of a single investor in the finite case as the number of investors goes to infinity (keeping total savings constant).

Interestingly, for any price level p_b, $e^{\dagger}_{\text{soph}} < e/p_b$. In contrast to the naive specification (3), as a marginal share of green investors switches from the green to the brown technology, they increase the demand for brown investments, triggering a higher brown price, and thus end up buying stricly less than $1/p_b$ units of brown investment. Put provocatingly, (higher) sophistication here provides investors with (more) excuses to save in the brown technology. When $p_b \geq c_b$, the relative difference between the sophisticated and naive perceptions of the induced externality is equal to:

$$\frac{\frac{e}{p_b} - e^{\dagger}_{\text{soph}}}{e^{\dagger}_{\text{soph}}} = \eta_b^{-1}$$

when $p_b \geq c_b$, and 0 otherwise. Hence, the naive perception (3) overestimates impact – equivalently, the sophisticated perception (4) provides an excuse to save in the brown technology – with a (relative) factor equal to η_b^{-1}, the inverse of the price-elasticity of brown capital.

Observation 1 (Sophistication and perceived induced externality). *With direct consequentialism, the higher the investors' sophistication, the lower their perceived induced externality at any given prices and capital stocks. As a consequence, the higher the investors' sophistication, the lower the green premium and the higher the aggregate pollution.*

Moreover, with a mild additional assumption on the investment supply functions, the same comparative statics with respect to the unit externality e and distribution of willingnesses-to-do-good F (in the first-order-stochastic-dominance sense) hold with sophisticated investors as with naive ones. By contrast, the comparative statics with respect to capital stocks and investment supply function can be more ambiguous (see Appendix A for details).

2.4 Welfare

To draw a welfare analysis, we detail the production side of our competitive economy. We assume for each technology the existence of a representative producer (with no ethical concerns). For $s \in \{b, g\}$, given equilibrium prices p_s, the representative producer for technology

By market clearing and a Taylor expansion, this expression is equivalent to (4).

s therefore chooses its investment i_s by solving

$$\max_{i_s \geq 0} \quad p_s i_s - C_s(i_s)$$

We assume strictly increasing and convex costs, with $C_g'(0)$, $C_b'(0) > 0$. The solution is given by the first-order condition $C_s'(i_s) = p_s$ for any $p_s \geq C_s'(0) = c_s$, and by $i_s = 0$ whenever $p_s < C_s'(0) = c_s$.

We consider a utilitarian and materialistic welfare objective:

$$W \equiv (1-e)(k_b + i_b) + (k_g + i_g) - C_b(i_b) - C_g(i_g)$$

Let us assume that $c_g \leq 1$ (positive green investment at the optimum). Two cases arise. If $1 - e < c_b$ (no brown investment other than the existing stock at the optimum), the first-best prices satisfy:

$$p_b \leq c_b, \qquad \text{and} \qquad p_g = 1.$$

If $1 - e \geq c_b$ (positive brown investment at the optimum), the first-best prices satisfy:

$$p_b = 1 - e, \qquad \text{and} \qquad p_g = 1.$$

For simplicity, we assume that the investment market clears at the first-best prices. Specifically, we assume that

$$(1-e)[k_b + (C_b')^{-1}(1-e)] + [k_g + (C_g')^{-1}(1)] = 1, \tag{5}$$

thus uniquely pinning down the brown price subject to market clearing in the case $1 - e < c_b$.

Absent investors' environmental concerns, *laissez-faire* always yields under-investment in the green technology with respect to the welfare optimum, and if $(k_b + k_g)\min(c_b, c_g) < 1$ (positive investment under *laissez-faire*), over-investment in the brown technology. With environmental concerns, there can be either over- or under-investment in each technology with respect to the welfare optimum.

Therefore, a principal with welfare objective W may contemplate policy interventions, which we study next.

3 Policy

3.1 Taxation

Let us consider a tax τ_b on returns on brown savings.[18] We successively consider two utility specifications, depending on whether investors internalize the impact of their saving decision on tax proceeds.

Narrow internalization (environmental externality alone). Investors only internalize the induced environmental externality $e^\dagger$ and ignore their impact on tax proceeds. The investors' segmentation thus writes as

$$e^\dagger v^* - \left(\frac{1-\tau_b}{p_b} - \frac{1}{p_g} \right) = 0, \tag{6}$$

whereas Equations (2)-(3) are unchanged.

A tax on brown returns (or a subsidy on green returns) lowers the brown equilibrium price p_b, the level of brown investment i_b and the fraction of brown investors $F(v^*)$ while increasing the green price p_g, and the level of green investment i_g. While this implies that taxes on brown returns and/or subsidies on green returns increase the pre-tax green premium, the comparative statics of the tax-adjusted green premium are ambiguous.

Broad internalization (the "tax excuse"). Investors now internalize both the induced externality $e^\dagger$ and their marginal induced impact on tax proceeds $t^\dagger$, where their (naive) perception of $t^\dagger$ is given $t^\dagger = \tau_b/p_b$ for any $p_b \geq 0$ (see below for the expression of $t^\dagger$ with sophisticated investors).[19] The equilibrium cut-off is given by:

$$(e^\dagger - t^\dagger)v^* - \left(\frac{1-\tau_b}{p_b} - \frac{1}{p_g} \right) = 0$$

Comparative statics may differ from those with narrow internalization.[20]

We assume there is no cost of collecting taxes. We focus on the case where $c_g \leq 1$ (positive green investment at the optimum). To avoid technical issues related to the shape of the

[18]Considering subsidies τ_g on the returns of green savings would lead to the same conclusions.

[19]Hence, with naive investors,

$$t^\dagger = \frac{\tau_b}{e} e^\dagger \mathbf{1}_{p_b \geq c_b} + \frac{\tau_b}{p_b}\left(1 - \mathbf{1}_{p_b \geq c_b}\right)$$

[20]As long as $p_b(\tau_b) \geq c_b$, broad internalization yields the same comparative statics as narrow internalization for a tax level $\tau_b < e$ if and only if $(1-e)\big/p_b(\tau_b) - 1\big/p_g(\tau_b) < 0$.

distribution of types F and cost functions C_b, C_g, we focus in the next Proposition on the implementation of the first-best prices.[21]

Proposition 2 (**Consequentialism and Pigouvian taxation**). *Assume* $c_g \leq 1$*, and that market clearing at first-best prices (5) holds. If* $1 - e \geq c_b$*, then*

(i) With narrow internalization, any tax rate that implements the first-best prices (if any) is strictly lower than the Pigouvian tax ($\tau_b < e$*), due to the investors' environmental concerns, unmitigated by their impact on tax proceeds. The green premium is strictly positive.*

(ii) With broad internalization, the Pigouvian tax ($\tau_b = e$*) implements the first-best prices. The green premium is nil.*

By contrast, if $1 - e < c_b$*, then*

(i) With narrow internalization, the Pigouvian tax ($\tau_b = e$*) implements the first-best prices. The green premium is nil.*

(ii) With broad internalization, any tax rate that implements the first-best prices (if any) is strictly higher than the Pigouvian tax ($\tau_b > e$*), due to the investors' "tax excuse". The green premium is strictly negative.*

As intuitive, with broad internalization (and thus the "tax excuse"), the taxes that implement the first-best prices (if any) are strictly higher than with narrow internalization (without the "tax excuse").

When $1-e \geq c_b$ (positive brown investment at the first-best prices), under broad internalization, the standard Pigouvian tax achieves the first-best, while making investors indifferent between saving in the green or brown technologies, as the negative environmental externality of the latter is exactly compensated by its positive impact on tax proceeds.[22] When $1-e < c_b$ (no brown investment at the first-best prices), the induced environmental externality $e^\dagger$ at first-best prices is nil, whereas the perceived contribution to tax proceeds $t^\dagger$ remains strictly positive. As a consequence, a necessary condition for a tax rate to implement the first-best prices with broad internalization is that it be strictly higher than the Pigouvian tax.

[21] Note that in case $1 - e \geq c_b$ with narrow internalization or in case $1 - e < c_b$ with broad internalization, a tax that implements the first-best prices may not exist. See Appendix G.2 for details.

[22] Moreover, except in the non-generic case $F(1) = (1-e)[k_b + (C_b')^{-1}(1-e)]$, the Pigouvian tax is *uniquely* optimal whenever it is optimal.

Sophistication and Pigouvian taxation. How much of the above results stem from the investors' naiveté in assessing their environmental and tax-collection externality? In fact, with sophisticated investors rather than naive ones, i.e. with investors taking into account the impact of their investment decision on equilibrium prices (see Section 2.3), the relation between the environmental externality $e^{\dagger}_{\text{soph}}$ and the tax-collection externality $t^{\dagger}_{\text{soph}}$ is given by:

$$t^{\dagger}_{\text{soph}} = \frac{d(\tau_b(k_b + i_b))}{f(v^*)dv^*} = \frac{\tau_b di_b}{f(v^*)dv^*}\mathbf{1}_{p_b \geq c_b} = \frac{\tau_b}{e}e^{\dagger}_{\text{soph}}$$

for all $p_b \geq 0$. In particular, for $p_b < c_b$, an investor's marginal impact on tax proceeds $t^{\dagger}_{\text{soph}}$ is nil. For, the brown market clearing condition yields that the share of brown investors is then equal to $p_b k_b$, and thus implies that the total amount collected from the tax is equal to $\tau_b k_b$, which does not depend on p_b.

Lemma 2 (**Sophistication, broad internalization and Pigouvian taxation**). *Assume $c_g \leq 1$, and that market clearing at first-best prices (5) holds. Assume that investors are sophisticated, thus with $e^{\dagger}_{soph}$ given by (3) and $t^{\dagger}_{soph} = (\tau_b/e)e^{\dagger}_{soph}$. Then, with broad internalization, the Pigouvian tax ($\tau_b = e$) implements the first-best prices, regardless of whether $1 - e \geq c_b$.*

3.2 Divestment

How effective, if at all, are (mandated) divestment strategies when investors abide by direct consequentialism? Let us assume that a fraction $\alpha \in (0, 1)$ of investment is channeled through a public entity (e.g. a public pension fund) whose portfolio composition can be mandated by a public decision-maker, while the complementary share $(1 - \alpha)$ accrues from private investors who can invest as they like. The public entity's endowment α is fixed, and may be managed by different sub-entities, which we refer to as "public investors". A public investor's environmental sensitivity, which stems from its investment mandate, is embodied in its type. We assume the distributions of types among public and private investors are given respectively by the smooth distributions G and F, which have full support on $\mathbb{R}_+$.

We examine the impact of the public entity divesting from the brown technology. Divestment is decided by the public decision-maker who has authority over the public entity. While under *laissez-faire*, each public investor invests as it likes given its environmental sensitivity (i.e. has flexible portfolio choice), under (mandated) divestment, each public investor must invest in the green technology.[23,24]

As a consequence, in the absence of (mandated) divestment, the equilibrium is given by the cut-off types among public and private investors, respectively denoted by v^*_{pub} and v^*_{priv}, together with prices p_b^{nd}, p_g^{nd} and perceived induced externality $e^\dagger_{\text{nd}}$ satisfying the investors' segmentation conditions:

$$e^\dagger_{\text{nd}} v^*_{pub} = e^\dagger_{\text{nd}} v^*_{priv} = \frac{1}{p_b^{\text{nd}}} - \frac{1}{p_g^{\text{nd}}},$$

where we have used that with direct consequentialism, both public and private investors face the same perceived induced externality:

$$e^\dagger_{\text{nd}} = \frac{e}{p_b^{\text{nd}}} \mathbf{1}_{\{p_b^{\text{nd}} \geq c_b\}}.$$

Lastly, the market clearing conditions write as:

$$\alpha G(v^*_{pub}) + (1-\alpha) F(v^*_{priv}) = (k_b + h_b(p_b^{\text{nd}}))p_b^{\text{nd}} = 1 - (k_g + h_g(p_g^{\text{nd}}))p_g^{\text{nd}}$$

By contrast, if the public entity divests from the brown technology, the equilibrium is given by the cut-off type among private investors, denoted by v^*, together with prices p_b^{d}, p_g^{d} and perceived induced externality $e^\dagger_{\text{d}}$ satisfying the investors' segmentation condition:

$$e^\dagger_{\text{d}} v^* = \frac{1}{p_b^{\text{d}}} - \frac{1}{p_g^{\text{d}}},$$

while the market clearing conditions now write as:

$$(1-\alpha) F(v^*) = (k_b + h_b(p_b^{\text{d}}))p_b^{\text{d}} = 1 - (k_g + h_g(p_g^{\text{d}}))p_g^{\text{d}}$$

As the public sector orients all its investments towards the green technology, it deteriorates the financial returns on green investments while improving those on brown investments. Hence, additional private investors (with types $v \geq v^*_{priv}$) turn to the brown technology instead of the green one (and thus $v^* > v^*_{priv}$): the impact of divestment is marred by a (partial) leakage. Nonetheless, this increase in private brown investments does not offset the decrease in public

[23] This may alternatively be interpreted as setting the investment mandate of all public sub-entities to investing in the green technology, i.e. changing their type to $+\infty$. More on this below.

[24] In a more complex environment with uncertainty on the distribution of environmental awarenesses (types) and a superiorly informed government, public policies can have an expressive role as private investors may try to infer the social norm from the public entities' investment mandate. However, in contrast to Bénabou and Tirole (2011), markets would limit the government's ability to use expressive public policies as investors may recover the social norm by observing the equilibrium prices, which derive from market clearing.

brown investments: overall the level of brown investments decreases when the public sector divests from the brown technology.

By contrast, consider the case in which prior to divestment there was no additional brown investment (with all brown savings going to existing capital stocks and $p_b \leq c_b$). In such a case, divestment has no impact on the externality level as the total induced externality remains equal to ek_b.

Proposition 3 (Divestment). *(i) Starting from a strictly positive level of brown investment, i.e. if $[k_b+k_g+h_g(c_b)]c_b \leq 1$, public divestment generates lower aggregate pollution and a higher green premium.*

(ii) By contrast, starting from zero brown investment, i.e. if $[k_b + k_g + h_g(c_b)]c_b > 1$, public divestment has no impact on aggregate pollution nor on the green premium (which remains nil).

In both cases, divestors reap lower financial returns than non-divestors, and brown divestment lowers the financial returns of private green investors.

Nonetheless, a lower externality level due to divestment may come at too high a cost in terms of lower production of the consumption good. Let us assume that the (utilitarian) welfare objective W is quasi-concave in brown investment i_b. Then, starting from an initial state with excessive brown investment and insufficient green investment with respect to the first-best levels, a marginal divestment increases social welfare.[25]

3.3 Investing for change

The same technology can in practice often be operated in several ways, yielding different levels of externalities. What if via shareholder activism, investors could make the brown technology cleaner? Would direct-consequentialist investors implement change? We characterize conditions under which, if shareholders are pivotal, change can be implemented. We show that the moral pecking order can vary, and that investing in the brown technology in order to subsequently clean it can be the "morally right" decision.[26]

To focus on investors' moral choices, we abstract away from the issues of collective decision-making. We thus assume that production is divisible, and that each investor has full decision rights over the production units she has bought,[27] and that brown production units can be

[25] On the opposite, starting from an initial state with insufficient brown investment and excessive green investment relative to the first-best levels, a marginal divestment is welfare-detrimental.

[26] We refer to a saving decision as being the "morally right" one if it is followed by the highest types in equilibrium.

[27] Each investor in technology s buys $1/p_s$ production units. We thus assume that each investor in technology

made less polluting, i.e. reach a unit externality $\tilde{e} < e$. Cleaning the brown technology entails a productivity loss $\delta > 0$ per unit of investment, and thus one unit of cleaned brown investment delivers $1 - \delta$ unit of the consumption good. Timing is as follows:

(1) investors save either in the green or in the brown technology,[28]

(2) brown investors decide whether or not to implement change (i.e. clean their production units),

(3) production takes place and payoffs accrue.

We show that the moral pecking order between "investing in the green technology" and "investing in the brown technology and cleaning it" depends on whether additional brown investments are profitable. Indeed, whenever $[k_b + k_g + h_g(c_b)]c_b \leq 1$, the equilibrium is described by two cut-off types $\underline{v} \leq \overline{v}$ such that types $v \geq \overline{v}$ invest in the green technology, types $v \in [\underline{v}, \overline{v})$ invest in the brown technology and clean it, and types $v < \underline{v}$ invest in the brown technology and keep it dirty. If $\underline{v} < \overline{v}$, the two cut-offs are thus given by the investors' segmentation conditions:[29]

$$\begin{cases} \dfrac{\tilde{e}}{p_b}\overline{v} - \left(\dfrac{1-\delta}{p_b} - \dfrac{1}{p_g}\right) = 0 \\ \dfrac{(e-\tilde{e})}{p_b}\underline{v} - \dfrac{\delta}{p_b} = 0, \end{cases}$$

together with the market clearing conditions:

$$F(\overline{v}) = (k_b + h_b(p_b))p_b, \qquad \text{and} \qquad 1 - F(\overline{v}) = (k_g + h_g(p_g))p_g,$$

whereas if $\underline{v} = \overline{v}$, the cut-off is given by v^* satisfying (1)-(2) as in the baseline setting. The cut-offs are unique.

By constrast, if $[k_b + k_g + h_g(c_b)]c_b > 1$ (and thus $p_b < c_b$), an equilibrium is described by a cut-off $\tilde{v}$ such that types $v \geq \tilde{v}$ invest in the brown technology and clean it, while types

s owns a firm of "size" $1/p_s$, of which she is the sole decision-maker. Hence, we focus on the polar case in which each investor is *pivotal*. From a direct consequentialist perspective, the absence of pivotality would annihilate the moral incentives to avoid holding brown shares. By contrast, this moral pressure would persist even in the absence of pivotality under both alternative moral criteria studied in Section 4.

[28]As a consequence, since all saving decisions take place at step 1, the prices of both technologies are fixed given the investors' decisions at step 1.

[29]The expressions for the perceived induced externalities stem from the fact that from a direct consequentialist perspective, investors should consider the next most attractive investment option as a counterfactual when assessing the induced externality of their saving decision. Here, each investor in technology $s \in \{b, g\}$ owns a firm producing $1/p_s$ unit of the consumption good. An investor with type $\underline{v}$ is indifferent between cleaning or not her brown firm: since her firm produces $1/p_b$ units of the consumption good, the total externality generated by the firm is equal to $\tilde{e}/p_b$ if cleaned, and e/p_b otherwise.

$v < \tilde{v}$ invest indifferently in the green or in the brown technology (leaving the latter dirty). The cut-off $\tilde{v}$ is given by the investors' segmentation condition:

$$\frac{(e - \tilde{e})}{p_b}\tilde{v} - \frac{\delta}{p_b} = 0,$$

together with the price parity condition $p_g = p_b$ $(< c_b)$. The cut-off $\tilde{v}$ is unique.

Proposition 4 (**Investing for change**). *The pecking order induced by direct consequentialism is given by:*

(i) If $[k_b + k_g + h_g(c_b)]c_b \leq 1$, then: 1^o investing in the green technology (highest types), 2^o investing in the brown technology and cleaning it (intermediate types), 3^o investing in brown technology and not cleaning it (lowest types);

(ii) By contrast, if $[k_b + k_g + h_g(c_b)]c_b > 1$, then: 1^o investing in the brown technology and cleaning it (highest types), 2^o (ex aequo) investing in the green technology or investing in brown technology (lower types).

Claim (ii) may be the less intuitive part of Proposition 4. It states that when installed capital stocks are sufficiently high so that there is no additional investment in the brown technology, the "morally right" strategy is to invest in the latter and clean the brown stocks. Indeed, as the existing brown capital stocks are used for production anyway, the replacement argument fully plays: brown pollution will happen, generating a unit social cost e unless the technology is cleaned, in which case the unit social cost will be lowered to $\tilde{e}$. Hence, from a direct consequentialist perspective, investing in the green technology or in uncleaned brown installed stocks is equivalent in terms of induced externality. In contrast with these two investment strategies, investing in cleaned brown stocks reduces the externality level. The moral pecking order of claim (ii) follows. On the opposite, whenever additional brown investments are profitable, investing in the green technology remains the most moral thing to do as green investments remain less polluting than (additional) cleaned brown investments.

4 Alternative moral criteria

Let us now examine two alternative moral criteria: *rule consequentialism* and *shared responsibility*. Although both differ from direct consequentialism, they remain concerned with the impact of investment decisions.

	$[k_b + k_g + h_g(c_b)]c_b > 1$	$[k_b + k_g + h_g(c_b)]c_b \leq 1$
Externality $(-e(k_b + i_b))$	DC = RC = SR	RC < DC ≤ SR
Green premium (γ)	DC < RC < SR	RC < DC ≤ SR

Table 1: Comparison of direct consequentialism (DC), rule consequentialism (RC) and shared responsibility (SR).[32]

Rule consequentialism: What if everyone did the same? When making their saving decision, investors now take into consideration the induced externality they would generate if everyone mimicked their behaviour and invested in the brown technology. We denote by $e^\dagger_{rc}$ the resulting perceived induced externality.[30] We focus on installed brown capital stocks such that $k_b c_b < 1$. As a consequence, $e^\dagger_{rc}$ is strictly positive. The perceived induced externality $e^\dagger_{rc}$ may alternatively be interpreted as the (psychological) cost of infringing a moral imperative.

Comparison with direct consequentialism. Let p_b be the brown equilibrium price with direct consequentialism. Then, the perceived induced externalities with rule consequentialism and direct consequentialism are such that $e^\dagger_{rc} < e^\dagger$ if $[k_b + k_g + h_g(c_b)]c_b \leq 1$, and $e^\dagger_{rc} > e^\dagger$ if $[k_b + k_g + h_g(c_b)]c_b > 1$. Hence, whenever brown investments are profitable with direct consequentialism ($[k_b + k_g + h_g(c_b)]c_b \leq 1$), rule consequentialism leads to an equilibrium with a higher level of brown investment, a lower level of green investment and a higher externality level.[31]

Shared responsibility: Average impact of similar investments. While with direct consequentialism, investors care about the marginal impact of their saving decision, we now investigate the case in which they care about the average impact of similar investments. A (non-deontological) rationale for such a moral criterion may stem from the marginal contribution being hardly measurable, and/or more behavioural motives (such as group feelings, etc.).

With shared responsibility, investors feel responsible for the aggregate level of externalities

[30] Namely, letting p^*_b be such that $p^*_b(k_b + i_b(p^*_b)) = 1$, we have that $e^\dagger_{rc} = e/p^*_b = e(k_b + i_b(p^*_b))$.

[31] Intuitively, a rule consequentialist investor considers the hypothetical world in which all investors would take the same decision as her, and then computes her individual impact on the externality level in such a hypothetical world. In a world where all savers invest in the brown technology, the brown price is high, and thus each investor can only afford a small number of investment units, thus being individually responsible for a small externality.

[32] Whenever $[k_b + k_g + h_g(c_b)]c_b \leq 1$, inequalities between direct consequentialism and shared responsibility hold with equality with the naive perception of the induced externality (3), and are strict with the sophisticated one (4).

divided by the share of "wrongdoers": the perceived induced externality thus writes as

$$e_{sr}^{\dagger} = e\frac{k_b + i_b}{F(v^*)} = \frac{e}{p_b}$$

for all $p_b \geq 0$, where the second equality derives from the brown market clearing condition.

Comparison with direct consequentialism. With shared responsibility, the perceived induced externality coincides with the naive direct consequentialist perception (3) whenever $[k_b + k_g + h_g(c_b)]c_b \leq 1$, and is strictly higher otherwise. Hence, shared responsibility leads to an equilibrium with a (weakly) lower level of brown investment, a higher level of green investment, a higher green premium and lower aggregate pollution. Direct consequentialism may thus appear as a *weaker* moral criterion than shared responsibility. Interestingly, this dominance strenghtens with the sophisticated specification for the perceived induced externality under direct consequentialism (see Section 2.3), as the latter is always strictly below the one with shared responsibility.

Lastly, let us consider taxation. With shared responsibility and broad internalization, the optimal tax rate is the Pigouvian tax ($\tau_b = e$), regardless of whether $1 - e \geq c_b$.[33]

Proposition 5 (**Rule consequentialism and shared responsibility**). *(i) Shared responsibility yields a lower aggregate pollution and a higher green premium than direct consequentialism, strictly so if investors are sophisticated. By contrast, rule consequentialism yields a lower pollution and a higher green premium than direct consequentialism if and only if* $[k_b + k_g + h_g(c_b)]c_b \leq 1$.

(ii) When $[k_b + k_g + h_g(c_b)]c_b \leq 1$*, the insights of Propositions 3 and 4 hold with shared moral responsibility as well as with rule consequentialism. By contrast, when* $[k_b + k_g + h_g(c_b)]c_b \leq 1$*, the neutrality results and setting-dependence of the moral pecking order fail (investing in the green technology being always strictly more "moral" than investing in the brown technology).*

5 Extensions

We briefly consider four extensions, referring to the Appendix for details.

[33]With shared responsibility and narrow internalization, any tax rate that implements the first-best prices (if any) is strictly lower than the Pigouvian tax, regardless of whether $1 - e \geq c_b$.

Image concerns. A large body of empirical evidence emphasizes that image concerns are major drivers of moral behaviour.[34] Hence, suppose that in addition to intrinsic motivation and monetary incentives, the agents' saving decision has a third driver: image concerns. Does this third driver affect the previous results? In particular, does it alter the design of optimal policies – e.g., as they may induce "greenwashing" incentives?

Specifically, let us assume that the agents have image concerns that are proportional to the perception of the induced externality $e^\dagger$, i.e., higher-externalities actions are more salient. The social image payoff associated with saving decision $a \in \{0, 1\}$ writes as $e^\dagger \mu \mathbb{E}[v|a]$, where $\mu \in \mathbb{R}_+$ is the strength of image concerns. Consequently, a type-v investor solves:

$$\max \left\{ e^\dagger (v + \mu \mathbb{E}[v|a = 1]) - \left(\frac{1}{p_b} - \frac{1}{p_g} \right) \; ; \; e^\dagger \mu \mathbb{E}[v|a = 0] \right\}$$

Under a standard additional condition (see Appendix B for details), all our previous results still hold if agents have image concerns. Furthermore, a higher intensity of image concerns, μ, induces a higher green premium.

Consumers with ethical preferences. Investors may not be the only ones to exhibit moral concerns: consumers' choices may also be ethically driven. Hence, some consumers may derive an additional utility from "doing good", e.g. from relying on energy produced from renewables instead of fossil fuels, from buying fair-trade and organic rather than standard food, or more generally goods from socially responsible firms or technologies. Appendix C thus departs from the assumption that one unit of the consumption good is identically valued irrespectively of its production technology. We consider ethically-motivated consumers with an intrinsic motivation for consuming goods produced with the green technology. The prices of consumption goods are equilibrium determined.

We find that (i) green consumers face a positive green premium, and that (ii) although they induce a lower green premium for green investors, the latter remains positive. In addition, brown divestment strategies now make green consumers better off and brown consumers worse off.

Decommissioning of installed capital stocks. Let us briefly discuss how decommissioning affects the above analysis, referring to Appendix D for details. Suppose existing capital

[34] See, e.g., Bénabou and Tirole (2006) for a literature review. In addition to making the model more realistic, introducing image concerns may pave the way for future work on norms-based interventions – e.g., via public disclosure of consumption and investment patterns –, and on *moral narratives*, which we briefly discuss in Section 7.

stocks can be decommissioned at some (strictly positive) cost – e.g. the cost of dismantling a plant and cleaning the area –, and thus that there exists an induced externality for operating existing capital stocks which would otherwise be decommissioned. Then, similar comparative statics to the ones of Proposition 1 obtain when the producers are able to decommission the existing capital stocks.

Moreover, public brown divestment has no impact on the externality level whenever the magnitude of divestment is not large enough to make decommissioning profitable for brown producers. In addition, when decommissioning is possible, the moral pecking order of Proposition 4-(ii) – that is, cleaned brown technology first, then indifferently green and uncleaned brown second – only obtains on the range of brown equilibrium prices not low enough to induce decommissioning of brown capital stocks. For prices sufficiently low to induce the latter, the pecking order of Proposition 4-(i) obtains.

Minimum investment return. Suppose that each investor must achieve a (financial) return no lower than $l > 0$ on its saving decision – e.g. to face an (idiosyncratic) liquidity shock, or because their investment mandate requires them not to underperform with respect to a given benchmark. Let us then allow for non-binary saving decisions, with $a \in [0, 1]$ denoting the share of the investor's wealth saved in the green technology. Hence, any investor faces the minimum-return constraint:

$$\frac{1-a}{p_b} + \frac{a}{p_g} \geq l.$$

Let us look for an equilibrium in which prices satisfy: $1/p_g < l < 1/p_b$, and low types fully invest in the brown technology ($a = 0$), while high types invest a share $\overline{a} \in (0, 1)$ of their wealth in the green technology, where

$$\overline{a} = \frac{\frac{1}{p_b} - l}{\frac{1}{p_b} - \frac{1}{p_g}},$$

i.e. $\overline{a}$ is the largest share they can save in the green technology while still meeting the minimum-return constraint.[35]

[35] We assume that such an equilibrium exists (for some value of l). With perceived induced externality $e^\dagger$ (per unit of brown investment), the investors' segmentation condition thus writes as

$$(1-\overline{a})e^\dagger v^* + \frac{1-\overline{a}}{p_g} + \frac{\overline{a}}{p_b} = \frac{1}{p_b},$$

i.e. equivalently as

$$e^\dagger v^* + \frac{1}{p_g} - \frac{1}{p_b} = 0.$$

Letting v^* be the cutoff type, the market clearing conditions write as

$$p_b[k_b + i_b] = 1 - p_g[k_g + i_g] = F(v^*) + (1 - \overline{a})[1 - F(v^*)].$$

Let us focus on the case $p_b \geq c_b$ (positive brown investment). The naive perception of the induced externality of saving one unit of numéraire in the brown technology is thus equal to e/p_b, as before. Similarly, the sophisticated perception of the induced externality is given (as before) by

$$e^\dagger_{\text{soph}} = \frac{edi_b}{f(v^*)dv^*},$$

yet now, the market clearing conditions and the definition of $\overline{a}$ yield that[36]

$$e^\dagger_{\text{soph}} = \frac{e}{p_b(1+\eta_b^{-1})} \cdot \frac{1}{1 + \frac{1-F(v^*)}{\left(\frac{1}{p_b}-\frac{1}{p_g}\right)^2}\left[\left(\frac{1}{p_b} - l\right)\frac{1}{p_g^2}\frac{1}{p_g h_g'(p_g)(1+\eta_g^{-1})} - \left(l - \frac{1}{p_g}\right)\frac{1}{p_b^2}\frac{1}{p_b h_b'(p_b)(1+\eta_b^{-1})}\right]}$$

The sign of the term between brackets in the denominator is ambiguous in general – which suggests that in complex environments, sophistication may lead to a *higher* perceived externality than naiveté (in contrast to our simple setting and Observation 1). Intuitively, this ambiguity stems from two contrasting effects:

(i) *Substitution* (with magnitude $(1/p_b - l)$): buying a brown share instead of a green one leads to a lower green price, and thus higher green returns, and consequently, all else being equal, investors willing to do good can substitute some brown shares with green ones, while still meeting the minimum-return constraint;

(ii) *Reinforcement* (with magnitude $(l - 1/p_g)$): buying a brown share instead of a green one leads to a higher brown price, and thus lower brown returns, and consequently, all else being equal, investors must buy (even) more brown shares to meet the minimum-return

[36] Indeed, letting for $s \in \{b, g\}$, η_s be the price-elasticity of s-capital, total differentiation of the market-clearing condition yields:

$$p_b(1+\eta_b^{-1})di_b = f(v^*)dv^* - [1 - F(v^*)]d\overline{a},$$

while total differentiation of the definition of $\overline{a}$ yields:

$$d\overline{a} = \frac{-\left(l - \frac{1}{p_g}\right)\frac{dp_b}{p_b^2} - \left(\frac{1}{p_b} - l\right)\frac{dp_g}{p_g^2}}{\left(\frac{1}{p_b} - \frac{1}{p_g}\right)^2} = \frac{\left(\frac{1}{p_b} - l\right)\frac{1}{p_g^2}\frac{p_b h_b'(p_b)(1+\eta_b^{-1})}{p_g h_g'(p_g)(1+\eta_g^{-1})} - \left(l - \frac{1}{p_g}\right)\frac{1}{p_b^2}}{\left(\frac{1}{p_b} - \frac{1}{p_g}\right)^2} dp_b.$$

constraint.

We leave for future work a detailed analysis of such environments.

6 Related literature

A rich literature has studied prosocial behaviours, which cannot be fully reviewed here. Bénabou and Tirole (2006, 2011) introduce the three-fold utility specification (intrinsic motivation, image concerns, material incentives) on which we build.[37] Our paper further builds on a series of works investigating corporate social responsibility,[38] investors' activism and how they may successfully mitigate a firm's negative externalities. While contributions to this literature are numerous and diverse, this paper's focus on the systemic implications of (several) moral criteria is, to the best of our knowledge, new. Notable contributions include Heinkel et al. (2001) who study the effect of exclusionary ethical investing on corporate behaviour in a risk-averse setting,[39] as well as Gollier and Pouget (2014) who show that a large activist investor internalizing the social impacts of her investments may generate positive abnormal returns by investing in non-responsible companies and turning them into responsible ones. More recently, Hart and Zingales (2017) ask the question of the appropriate objective function for a firm when shareholders have an intrinsic motivation for increasing the social good and externalities are not perfectly separable from production decisions, stressing that maximisation of shareholder welfare may differ from maximisation of market value.[40,41] In a related vein, Oehmke and Opp (2023) draw a model of corporate financing under agency frictions assuming that socially responsible investors care about the firm's externalities irrespectively of whether they are investors in the firm (which they show to be a necessary condition for those investors to impact the firm's behaviour), and derive an investment criterion to guide

[37] Empirical evidence for the relevance of such a utility specification in the specific context of financial investments is provided by Riedl and Smeets (2017) who find that SRI is driven both by social preferences and social signalling. In a related vein, Hartzmark and Sussman (2019) find evidence consistent with one's intrinsic motivation for doing good influencing one's expectations of sustainable fund performance.

[38] See e.g., Bénabou and Tirole (2010) and references therein.

[39] The authors find that a lack of risk-sharing among brown investors leads to lower stock prices for polluting firms. If a higher cost of capital more than overcomes the cost of reforming, then polluting firms become socially responsible because of exclusionary ethical investing.

[40] The authors assume a particular form of consequentialism: shareholders vote as if they were pivotal, yet only feel responsible for the outcome if their vote *was* pivotal.

[41] Relatedly, Morgan and Tumlinson (2019) draw a model in which shareholders who care about a public good optimally require the firm's management to sacrifice some profits to contribute to the public good, due to two efficiency gains: overcoming freeriding incentives that appear when contributions to the public good are decentralized, and direct control of production and thus externality levels.

scarce socially responsible capital from an impact-maximizing perspective.[42,43] Closely related to our setting, Green and Roth (2023) offer a rich framework of a competitive financial market in which commercial investors compete with socially-motivated investors who are either *naive* or *sophisticated* in assessing the impact of their investment decisions (the former failing to take into account the displacement effects on other investors induced by their own choices, the latter being concerned with the aggregate social output), investigating the equilibrium and optimal allocation of social capital. Hence, the environment of Green and Roth (2023) may be interpreted in the light of our setting as the case where investors with no social preferences coexist with both naive consequentialist investors and investors concerned with the aggregate externality level. The authors notably show that increasing profitability can have a greater social impact than focusing on direct social value creation. Also considering selfish, commercial investors competing with consequentialists (as well as with deontologists), Schmidt and Herweg (2021) examine the optimality of price or quantity regulations, showing that when the environmental tax is bounded away from the optimum (say, for political reasons), price regulation dominates quantity regulation in terms of common good supply and welfare. Lastly, in a slightly different setting, Hakenes and Schliephake (2023) investigate the optimal combination of socially responsible investment and consumption, showing in particular that for a given minimum acceptable utility, socially-motivated agents maximize their impact on the aggregate externality by cutting their consumption of the polluting good in the same proportion as their divesting from the polluting firm.[44] While many contributions to this literature have derived interesting insights by studying models featuring investors with different views on morality – governing their assessments of the goodness of their actions –, our paper investigates the benchmark case in which investors agree on a moral criterion, although with heterogeneous degrees of motivation for doing good.

As our paper is concerned with capital markets, it also connects to the literature on financial returns when investors may have prosocial preferences. Since the seminal work of Fama and French (2007) who establish an asset-pricing framework allowing for disagreement and tastes for assets as consumption goods, contributions related to socially responsible investments have

[42]Their index may be interpreted as obeying a consequentialist ethos as it highlights the importance of the counterfactual social costs that would arise in the absence of socially responsible investors.

[43]Relatedly, from a consequentialist perspective, Landier and Lovo (2023) derive an ESG fund's optimal policy (in terms of industry allocation and pollution limits imposed on portfolio companies) in a general equilibrium model with search frictions on capital markets and where investors refuse lower returns than their best investment alternatives.

[44]Another alley of research examines the impact on production technologies of consumers preferences: Aghion et al. (2023) develop and test a model of innovation whereby socially responsible consumers induce firms to escape competition by pursuing green innovation. On social responsibility and corporate governance, see also Tirole (2001), Magill et al. (2015) and Broccardo et al. (2022).

been numerous, pointing in two distinct directions: one evidencing a positive "green premium" for some non-socially-responsible stocks in the wake of Hong and Kacperczyk (2009) who look at the effects of social norms on markets, and find that sin stocks have higher expected returns than otherwise comparable stocks;[45] the other one emphasizing that green investments may constitute a profitable hedging strategy, as for instance Andersson et al. (2016) who build a dynamic investment strategy that allows long-term passive investors to hedge climate risk without sacrificing financial returns.[46,47] In this paper, we rule out uncertainty on future environmental externalities as well as regulation, and focus on the risk-adjusted returns in order to isolate the role of moral criteria. The premia we find are consistent with the empirical evidence.

Our modelling assumptions are motivated by a vast empirical literature. Several experimental studies have evidenced that investors do care about the externalities their investments may generate. Bartling et al. (2014) set a product market in which low-cost production creates a negative externality for third parties, but where alternative production with higher costs mitigates the externality, and find a persistent preference among many consumers and firms for avoiding negative social impact in the market, reflected both in the composition of product types and in a price premium for socially responsible products. Most directly related to our focus on direct consequentialism, Bonnefon et al. (2022) measure how shareholders value a firm's ethical actions and find in particular that (i) investors are willing to pay more for buying a share in a firm generating positive externalities, while symmetrically, a firm that makes profits by exercising a negative externality is valued less than a similar company with no externality, and that (ii) pivotality does not affect an investor's willingness to pay;[48] relatedly, Brodback et al. (2022) propose an experimental study of investors' willingness-to-pay for socially responsible assets, and find that (i) individuals attribute a positive value to social responsibility at an increasing rate, and furthermore that (ii) assets generating an extra-financial benefit when financial performance is bad suffer from a price discount.

Lastly, in examining the interplay of (utilitarian) ethics and economics, this paper follows

[45] In this line of research, see also El Ghoul et al. (2011), Bolton and Kacperczyk (2021) on the higher returns of stocks of companies generating negative externalities, as well as Baker et al. (2022) and Zerbib (2019) on the lower returns of green bonds, and Barber et al. (2021) on the lower returns of impact investments. On the theoretical side, Pástor et al. (2021) offer a model of SRI in which, due to the investors' intrinsic preference for holding green assets and due to these assets' hedging of climate risk, green (resp. brown) assets have negative (resp. positive) CAPM alphas in equilibrium, and yet unexpected shifts in investors' tastes for green assets may induce the latter to outperform brown assets.

[46] Indeed, in their paper, green investment features as a "free option on carbon": once carbon dioxide emissions are priced, or expected to be priced, the low-carbon index should start to outperform the benchmark.

[47] On the profitability of SRI, see also Hong et al. (2012) for an investigation of the direct vs. selection effects.

[48] Hence investors' behaviour may be inspired by a *rule* consequentialist moral criteria, rather than a strictly direct consequentialist one.

the seminal works of Harsanyi (1977a, 1977b). Most recently, Dewatripont and Tirole (2024) examine the impact of markets on ethical behaviour. In their environment, as in ours, market competition does not crowd out consequentialist ethics.

7 Alleys for future research

As the introduction covered the main insights of the paper, we conclude by evoking two areas that would benefit from future research.

Environmental narratives may govern the agents' feeling of "righteousness" and the allocation of social prestige by designating what (and who) is "right". Yet, narratives may not be consistent with a given moral criterion, or with full rationality. Trying to disentangle their interplay may thus yield interesting insights. Let us give two ilustrations. Building on the above analysis, a non-consequentialist narrative granting social prestige to green investors and inflicting social stigma on brown investors can increase pollution whenever the brown technology can be cleaned, and doing so is prescribed by direct consequentialism over investing in the green technology. By contrast, a naive narrative leading investors to take their induced externality as $e^{\dagger}$ (naive) rather than $e^{\dagger}_{\text{soph}}$ (sophisticated), thereby neglecting their impact on equilibrium prices, reduces aggregate pollution.

Lastly, the dynamics of the model may be worth investing. Indeed, as mentioned earlier, the green premium implies that brown investors receive higher returns than green investors. Does this imply that over time, investors with lower willingnesses to do good become increasingly richer, and those with higher willingnesses-to-do-good increasingly poorer? What would this imply for the dynamics of investments and technology development? Would consequentialism prescribe mixed intertemporal strategies, trading a higher externality today for (the ability to induce) a lower one tomorrow?

References

Adriani, F. and Sonderegger, S. (2019). "A Theory of Esteem-Based Peer Pressure". *Games and Economic Behavior*, 115(C):314–335.

Aghion, P., Bénabou, R., Martin, R., and Roulet, A. (2023). "Environmental Preferences and Technological Choices: Is Market Competition Clean or Dirty?". *American Economic Review: Insights*, 5(1):1–20.

Andersson, M., Bolton, P., and Samama, F. (2016). "Hedging Climate Risk". *Financial Analysts Journal*, 72:1–20.

Baker, M., Bergstresser, D., Serafeim, G., and Wurgler, J. (2022). "The Pricing and Ownership of US Green Bonds". *Annual Review of Financial Economics*, 14(1):415–437.

Barber, B. M., Morse, A., and Yasuda, A. (2021). "Impact Investing". *Journal of Financial Economics*, 139(1):162–185.

Bartling, B., Weber, R. A., and Yao, L. (2014). "Do Markets Erode Social Responsibility?". *The Quarterly Journal of Economics*, 130(1):219–266.

Bénabou, R. and Tirole, J. (2006). "Incentives and Prosocial Behavior". *American Economic Review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2010). "Individual and Corporate Social Responsibility". *Economica*, 77:1–19.

Bénabou, R. and Tirole, J. (2011). "Laws and Norms". *NBER Working Paper No. 17579.*

Bolton, P. and Kacperczyk, M. T. (2021). "Do Investors Care about Carbon Risk?". *Journal of Financial Economics*, 142:517–549.

Bonnefon, J.-F., Landier, A., Sastry, P. R., and Thesmar, D. (2022). "The Moral Preferences of Investors: Experimental Evidence". Working Paper 29647, National Bureau of Economic Research.

Broccardo, E., Hart, O., and Zingales, L. (2022). "Exit versus Voice". *Journal of Political Economy*, 130(12):3101–3145.

Brodback, D., Guenster, N., Pouget, S., and Wang, R. (2022). "The Valuation of Corporate Social Responsibility: A Willingness to Pay Experiment". Working paper.

Dewatripont, M. and Tirole, J. (forthcoming). "The Morality of Markets". *Journal of Political Economy.*

Ehlers, T., Mojon, B., and Packer, F. (2020). "Green Bonds and Carbon Emissions: Exploring the Case for a Rating System at the Firm Level". *BIS Quarterly Review*, September 2020.

El Ghoul, S., Guedhami, O., Kwok, C. C., and Mishra, D. (2011). "Does Corporate Social Responsibility Affect the Cost of Capital?". *Journal of Banking & Finance*, 35(9):2388–2406.

European Securities and Markets Authority (2023). "ESG Names and Claims in the EU Fund Industry". *ESMA TRV Risk Analysis*.

Fama, E. F. and French, K. (2007). "Disagreement, Tastes, and Asset Prices". *Journal of Financial Economics*, 83(3):667–689.

Gollier, C. and Pouget, S. (2014). "The "Washing Machine": Investment Strategies and Corporate Behavior with Socially Responsible Investors". TSE Working Papers 14-457, Toulouse School of Economics (TSE).

Green, D. and Roth, B. (2023). "The Allocation of Socially Responsible Capital". R&R at *Journal of Finance*.

Hakenes, H. and Schliephake, E. (2023). "Responsible Investment and Responsible Consumption". Working paper.

Harsanyi, J. C. (1977a). "Morality and the Theory of Rational Behavior". *Social Research*, 44(4):623–656.

Harsanyi, J. C. (1977b). "Rule Utilitarianism and Decision Theory". *Erkenntnis (1975-)*, 11(1):25–53.

Hart, O. and Zingales, L. (2017). "Companies Should Maximize Shareholder Welfare Not Market Value". *Journal of Law, Finance, and Accounting*, (2):247–274.

Hartzmark, S. M. and Sussman, A. B. (2019). "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows". *Journal of Finance*, 74(6):2789–2837.

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.

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., Kubik, J. D., and Scheinkman, J. A. (2012). "Financial Constraints on Corporate Goodness". Working Paper 18476, National Bureau of Economic Research.

Jewitt, I. (2004). "Notes on the Shape of Distributions". Working paper, Oxford University.

Landier, A. and Lovo, S. (2023). "ESG Investing: How to Optimize Impact?". HEC Paris Research Paper No. FIN-2020-1363.

Landier, A. and Nair, V. B. (2008). *Investing for Change: Profit from Responsible Investment.* Oxford University Press.

Magill, M., Quinzii, M., and Rochet, J.-C. (2015). "A Theory of the Stakeholder Corporation". *Econometrica*, 83(5):1685–1725.

Morgan, J. and Tumlinson, J. (2019). "Corporate Provision of Public Goods". *Management Science*, 65(10):4489–4504.

Oehmke, M. and Opp, M. (2023). "A Theory of Socially Responsible Investment". Conditionally accepted at *Review of Economic Studies.*

Pigou, A. C. (1920). *The Economics of Welfare.* London: Macmillan and Co.

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

Riedl, A. and Smeets, P. (2017). "Why Do Investors Hold Socially Responsible Mutual Funds?". *The Journal of Finance*, 72(6):2505–2550.

Schmidt, K. M. and Herweg, F. (2021). "Prices versus Quantities with Morally Concerned Consumers". Discussion paper.

Sinnott-Armstrong, W. (2023). "Consequentialism". In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Winter 2023 edition.

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

Zerbib, O. D. (2019). "The Effect of Pro-Environmental Preferences on Bond Prices: Evidence from Green Bonds". *Journal of Banking & Finance*, 98:39–60.

Appendix

A Sophisticated perception of the induced externality

Let the perceived induced externality $e^\dagger$ be given by (4) (see Section 2.3):

$$e^\dagger_{\text{soph}} = \begin{cases} \dfrac{eh_b'(p_b)}{k_b + h_b(p_b) + p_b h_b'(p_b)} > 0 & \text{if } p_b \geq c_b, \\ 0 & \text{otherwise.} \end{cases}$$

A sufficient yet not necessary condition for the results to hold is:

Assumption (A_2)**.** *The function h_b is such that $(p_b h_b')/(k_b + h_b)$ increases with $p_b \geq c_b$.*

With our microfoundations for investment supply, i.e. $h_b = (C_b')^{-1}$ with C_b the cost function of brown producers (strictly increasing and strictly convex), Assumption (A_2) is satisfied in particular for any cost function C_b with marginal cost function C_b' weakly concave (e.g., C_b quadratic).[49] Assumption (A_2) may equivalently be formulated as: $d(p_b e^\dagger_{\text{soph}})/dp_b \geq 0$ for any $p_b \geq c_b$: the perceived induced externality must not decrease too steeply as brown investments increase. Consequently, under Assumption (A_2), the equilibrium is uniquely given by (1)-(2)-(4).

We investigate in the rest of this section the changes induced by the sophisticated perception $e^\dagger_{\text{soph}}$. Proofs are given in the Appendix along the ones for the main specification.

Proposition A.1 (Comparative statics with perceived induced externality (4)). *Assume (A_2).*

(i) A higher (unit) externality e induces a lower brown price p_b, a higher green price p_g,

[49] Indeed, Assumption (A_2) writes as:

$$\frac{d}{dp_b}\left(\frac{p_b}{C_b''\big((C_b')^{-1}(p_b)\big)[k_b + p_b]}\right) \;\geq\; 0,$$

i.e. equivalently,

$$\frac{k_b C_b''\big((C_b')^{-1}(p_b)\big)^2 - C_b^{(3)}\big((C_b')^{-1}(p_b)\big)p_b[k_b + p_b]}{C_b''\big((C_b')^{-1}(p_b)\big)\Big(C_b''\big((C_b')^{-1}(p_b)\big)[k_b + p_b]\Big)^2} \;\geq\; 0.$$

and a higher green premium γ, while having an ambiguous impact on aggregate pollution $e(k_b + i_b)$.

(ii) *A higher (green or brown) capital stock k_s induces a lower green price p_g. A higher green capital stock k_g induces a lower brown price p_b.*

(iii) *A higher (green or brown) investment supply function h_s (considering homothetic variations) induces a lower technology-s price p_s. A higher green supply function h_g always induces a lower brown price p_b, while if k_b is sufficiently low, a higher brown supply function h_g induces a lower green price p_g. A higher greenn supply function h_g induces lower aggregate pollution, while if k_b is sufficiently small, a higher brown supply function h_b induces higher aggregate pollution.*

(iv) *A higher distribution F of willingnesses-to-do-good according to first-order stochastic dominance induces a lower brown price p_b, a higher green price p_g, a higher green premium γ and lower aggregate pollution $e(k_b + i_b)$.*

Taxation. Assuming (A_1) and (A_2), a tax on brown returns (or a subsidy on green returns) lowers the brown equilibrium price p_b, the level of brown investment i_b and the fraction of brown investors $F(v^*)$ while increasing the green price p_g, and the level of green investment i_g. While this result implies that taxes on brown returns and/or subsidies on green returns increase the pre-tax green premium, the comparative statics of the tax-adjusted green premium are ambiguous.

Furthermore, as noted in the text, with the sophisticated expression for the perceived induced externality, the optimality of the Pigouvian tax holds regardless of whether $1 - e \geq C_b'(0)$. Namely, we obtain the following result (see Appendix G.2 for the proof).

Proposition A.2. ***(Pigouvian taxation with sophisticated perceived induced externality)*** *Let the perceived induced externality $e^\dagger$ be given by (4). Assume (A_1) and (A_2). (i) With narrow internalization, the optimal tax rate is strictly lower than e due to the investors' intrinsic motivation and image concerns for contributing to the common good, unmitigated by their impact on tax proceeds. (ii) By contrast, with broad internalization, the optimal tax rate is given by the Pigouvian tax $\tau_b = e$, and achieves the first-best prices.*

Divestment. The assumption of direct consequentialism implies that both public and private investors face the same perceived induced externality[50], given by

$$e^\dagger_{\text{nd}} = \frac{eh_b'(p_b^{\text{nd}})}{k_b + h_b(p_b^{\text{nd}}) + p_b^{\text{nd}} h_b'(p_b^{\text{nd}})}$$

Assuming (A_2) in addition to (A_1), Proposition 3 holds with the perceived induced externality given by (4).

Investing for change. The uniqueness of cut-offs in the case $[k_b + k_g + h_g(c_b)]c_b \leq 1$ obtains assuming (A_2) in addition to (A_1). Proposition 4 obtains similarly.

B Details on image concerns

As mentioned in the text, with image concerns, type-v investor solves:

$$\max \left\{ e^\dagger (v + \mu \mathbb{E}[v|a=1]) - \left(\frac{1}{p_b} - \frac{1}{p_g}\right) \; ; \; e^\dagger \mu \mathbb{E}[v|a=0] \right\}$$

We denote by v^* the cut-off type below which agents save in the brown technology, and above which they save in the green one. Using the same notation as in Bénabou and Tirole (2011), the reputational incentive associated with investing in the green technology is given by:

$$e^\dagger \mu(\mathcal{M}^+(v^*) - \mathcal{M}^-(v^*)) = e^\dagger \mu \Delta(v^*)$$

where $\mathcal{M}^+(v^*) \equiv \mathbb{E}[v|v > v^*]$, $\mathcal{M}^-(v^*) \equiv \mathbb{E}[v|v < v^*]$, and $\Delta(v^*) \equiv \mathcal{M}^+(v^*) - \mathcal{M}^-(v^*)$. We assume that f is single peaked, and thus by Jewitt lemma,[51] Δ is U-shaped.

Assumption (A_1)**.** *For all* $v \geq 0$, $1 + \mu\Delta'(v) > 0$.

Assumption (A_1) guarantees the uniqueness of the cut-off type v^* for given prices and is standard in the literature on prosocial behaviour.[52]

[50] Indeed, letting $e^\dagger_{pub}$ and $e^\dagger_{priv}$ be respectively the perceived induced externality of public and private investors,

$$e^\dagger_{pub} = \left.\frac{edi_b}{\alpha f(v^*_{pub}) dv^*_{pub}}\right|_{dv^*_{priv}=0} = \left.\frac{edi_b}{(1-\alpha) f(v^*_{priv}) dv^*_{priv}}\right|_{dv^*_{pub}=0} = e^\dagger_{priv}$$

where the second inequality derives from the market clearing condition for the brown technology.

[51] We use the following result due to Jewitt (2004) in order to characterize the properties of Δ: If f is everywhere decreasing (increasing), then Δ is everywhere increasing (decreasing). If f has a unique interior maximum, then Δ has a unique interior minimum.

[52] Assumption (A_1) implies that the normalized (by the perceived induced externality), non-monetary incentive for the cut-off type to save in the green technology, that is the quantity $v^* + \mu\Delta(v^*)$, strictly increases with

Lemma B.1. *With Assumption* (A_1)*, there exists a unique cut-off type* v^**.*

Hence, the investors' segmentation is determined by

$$e^\dagger\left[v^* + \mu\Delta(v^*)\right] - \left(\frac{1}{p_b} - \frac{1}{p_g}\right) = 0 \tag{7}$$

The market clearing conditions of both technologies remain given by (2). The equilibrium prices are thus determined by the equilibrium in the equity market and the two market clearing conditions for brown and green portfolios. In particular, as F has no atom in 0, we will have that $v^* > 0$ in equilibrium.

Consequently, under Assumption (A_1), the equilibrium is uniquely given by (7)-(2)-(3).

Lemma B.2 (Green premium). *Assume* (A_1)*. The green premium* γ *is always positive, strictly so if and only if* $e^\dagger > 0$*, i.e. if and only if* $p_b \geq c_b$*. Conversely,* $e^\dagger = 0$ *if and only if* $p_g = p_b = p < c_b$ *with* $[k_b + k_g + h_g(p)]p \equiv 1$*.*

Proposition B.3 (Image concerns). *Assume* (A_1)*. Then, Lemma 1 and Proposition 1 hold (to the exception of the comparative statics with respect to* F *in Proposition 1). Moreover, the comparative statics of the model with respect to the strength of image concerns* μ *are the same as those with respect to the unit externality* e*.*

B.1 Taxation

Narrow internalization. Investors only internalize the induced externality $e^\dagger$ and ignore their impact on tax proceeds. The investors' segmentation thus writes as

$$e^\dagger\left[v^* + \mu\Delta(v^*)\right] - \left(\frac{1-\tau_b}{p_b} - \frac{1}{p_g}\right) = 0, \tag{8}$$

whereas Equations (2)-(3) are unchanged.

Assuming (A_1), a tax on brown returns (or a subsidy on green returns) lowers the brown equilibrium price p_b, the level of brown investment i_b and the fraction of brown investors $F(v^*)$ while increasing the green price p_g, and the level of green investment i_g. While this result

the cut-off type v^*. In particular, it rules out situations in which the (normalized) image incentive $\mu\Delta(v^*)$ decreases so steeply with the cut-off type, i.e. with the share of brown investors, that it offsets the variation in the (normalized) intrinsic motivation incentive. See Bénabou and Tirole (2006) for a detailed discussion.

implies that taxes on brown returns and/or subsidies on green returns increase the pre-tax green premium, the comparative statics of the tax-adjusted green premium are ambiguous.[53]

Broad internalization (The tax excuse). Investors now internalize both the induced externality $e^\dagger$ and their marginal induced impact on tax proceeds $t^\dagger$, where $t^\dagger = \tau_b/p_b$.[54] The equilibrium cut-off is given by:

$$(e^\dagger - t^\dagger)\left[v^* + \mu\Delta(v^*)\right] - \left(\frac{1-\tau_b}{p_b} - \frac{1}{p_g}\right) = 0.$$

Comparative statics may differ from those with narrow internalization.[55]

Proposition B.4 (Pigouvian taxation). *Assume* (A_1) *Assume* $c_g \leq 1$*, and that market clearing at first-best prices (5) holds. If* $1 - e \geq c_b$*, then*

(i) With narrow internalization, any tax rate that implements the first-best prices (if any) is strictly lower than the Pigouvian tax ($\tau_b < e$*), due to the investors' environmental concerns, unmitigated by their impact on tax proceeds. The green premium is strictly positive.*

(ii) With broad internalization, the Pigouvian tax ($\tau_b = e$*) implements the first-best prices. The green premium is nil.*

By contrast, if $1 - e < c_b$*, then*

(i) With narrow internalization, the Pigouvian tax ($\tau_b = e$*) implements the first-best prices. The green premium is nil.*

(ii) With broad internalization, any tax rate that implements the first-best prices (if any) is strictly higher than the Pigouvian tax ($\tau_b > e$*), due to the investors' "tax excuse". The green premium is strictly negative.*

[53] We note for further reference that while this ambiguity persists with shared responsibility, it disappears with rule consequentialism (see Section 4).

[54] Hence,

$$t^\dagger = \frac{\tau_b}{e} e^\dagger \mathbf{1}_{p_b \geq c_b} + \frac{\tau_b}{p_b}\left(1 - \mathbf{1}_{p_b \geq c_b}\right)$$

By contrast, as mentioned in the text, with sophisticated investors rather than naive ones, i.e. with investors taking into account the impact of their investment decision on equilibrium prices: $t^\dagger_{\text{soph}} = (\tau_b/e)e^\dagger_{\text{soph}}$ for all $p_b \geq 0$.

[55] As long as $p_b(\tau_b) \geq c_b$, broad internalization yields the same comparative statics as narrow internalization for a tax level $\tau_b < e$ if and only if $(1-e)\big/p_b(\tau_b) - 1\big/p_g(\tau_b) < 0$.

B.2 Divestment

Assume that the distributions F and G both satisfy Assumption (A_1). In the absence of (mandated) divestment, the equilibrium is given by the cut-off types among public and private investors, respectively denoted by v^*_{pub} and v^*_{priv}, together with prices p_b^{nd}, p_g^{nd} and perceived induced externality $e^\dagger_{\text{nd}}$ satisfying the investors' segmentation conditions:

$$e^\dagger_{\text{nd}}[v^*_{pub} + \mu\Delta_G(v^*_{pub})] = e^\dagger_{\text{nd}}[v^*_{priv} + \mu\Delta_F(v^*_{priv})] = \frac{1}{p_b^{\text{nd}}} - \frac{1}{p_g^{\text{nd}}},$$

where we have used that with direct consequentialism, both public and private investors face the same perceived induced externality:

$$e^\dagger_{\text{nd}} = \frac{e}{p_b^{\text{nd}}}\mathbf{1}_{\{p_b^{\text{nd}} \geq c_b\}}.$$

Lastly, the market clearing conditions write as:

$$\alpha G(v^*_{pub}) + (1-\alpha)F(v^*_{priv}) = (k_b + h_b(p_b^{\text{nd}}))p_b^{\text{nd}} = 1 - (k_g + h_g(p_g^{\text{nd}}))p_g^{\text{nd}}$$

By contrast, if the public entity divests from the brown technology, the equilibrium is given by the cut-off type among private investors, denoted by v^*, together with prices p_b^{d}, p_g^{d} and perceived induced externality $e^\dagger_{\text{d}}$ satisfying the investors' segmentation condition:

$$e^\dagger_{\text{d}}[v^* + \mu\Delta_F(v^*)] = \frac{1}{p_b^{\text{d}}} - \frac{1}{p_g^{\text{d}}},$$

while the market clearing conditions now write as:

$$(1-\alpha)F(v^*) = (k_b + h_b(p_b^{\text{d}}))p_b^{\text{d}} = 1 - (k_g + h_g(p_g^{\text{d}}))p_g^{\text{d}}$$

As the public sector orients all its investments towards the green technology, it deteriorates the financial returns on green investments while improving those on brown investments. Hence additional private investors (with types $v \geq v^*_{priv}$) turn to the brown technology instead of the green one (and thus $v^* > v^*_{priv}$): the impact of divestment is marred by a (partial) leakage. Nonetheless, we show that under our assumptions, this increase in private brown investments does not offset the decrease in public brown investments: overall the level of brown investments decreases when the public sector divests from the brown technology. The image incentive for private investors, i.e. the difference between the images associated to green and brown investments respectively, depends on the shape of the function Δ. In particular, if Δ is locally decreasing, the image incentive for private investors to save in the green technology rather than in the brown one decreases following public divestment.

By contrast, consider the case in which prior to divestment there was no additional brown investment (with all brown savings going to existing capital stocks) – e.g. because of very high environmental sensitivity e or strength of image concerns μ – and thus p_b was equal to c_b. In such a case, divestment has no impact on the externality level as the total induced externality remains equal to ek_b.

Proposition B.5 (Divestment). *Assume distributions F and G satisfy (A_1).*

(i) Starting from a strictly positive level of brown investment, i.e. if $[k_b + k_g + h_g(c_b)]c_b \leq 1$, public divestment generates lower aggregate pollution and a higher green premium.[56]

(ii) By contrast, starting from zero brown investment, i.e. if $[k_b + k_g + h_g(c_b)]c_b > 1$, public divestment has no impact on aggregate pollution nor on the green premium (which remains nil).

In both cases, divestors reap lower financial returns than non-divestors. Divestment lowers the financial returns of private green investors.

Nonetheless, a lower externality level due to divestment may come at too high a cost in terms of lower production of the consumption good. Let us assume distributions F and G satisfy (A_1), and that the welfare objective W is quasi-concave in brown investment i_b.[57] Then, starting from an initial state with excessive brown investment and insufficient green investment with respect to the first-best levels, a marginal divestment increases social welfare.[58]

B.3 Investing for change

We show that the moral pecking order between "investing in the green technology" and "investing in the brown technology and cleaning it" depends on whether additional brown investments are profitable. Indeed, whenever $[k_b + k_g + h_g(c_b)]c_b \leq 1$, the equilibrium is described by two cut-off types $\underline{v} \leq \overline{v}$ such that types $v \geq \overline{v}$ invest in the green technology, types $v \in [\underline{v}, \overline{v})$ invest in the brown technology and clean it, and types $v < \underline{v}$ invest in the brown technology and keep it dirty. If $\underline{v} < \overline{v}$, the two cut-offs are thus given by the investors'

[56]Moreover, if the distribution G of public investors' types is everywhere decreasing, then any increase in the strength of their image concerns μ_{pub} yields a lower externality level.

[57]A sufficient condition for this to hold is that both cost functions C_b and C_g be convex, with (weakly) positive third-derivatives $C_b^{(3)}$ and $C_g^{(3)}$.

[58]On the opposite, starting from an initial state with insufficient brown investment and excessive green investment relative to the first-best levels, a marginal divestment is welfare-detrimental.

segmentation conditions:[59]

$$\begin{cases} \dfrac{\tilde{e}}{p_b}[\overline{v}+\mu(\mathcal{M}^+(\overline{v})-\mathbb{E}[v|\underline{v}\leq v<\overline{v}])]-\left(\dfrac{1-\delta}{p_b}-\dfrac{1}{p_g}\right)=0 \\ \dfrac{(e-\tilde{e})}{p_b}[\underline{v}+\mu(\mathbb{E}[v|\underline{v}\leq v<\overline{v}]-\mathcal{M}^-(\underline{v}))]-\dfrac{\delta}{p_b}=0, \end{cases}$$

together with the market clearing conditions:

$$F(\overline{v})=(k_b+h_b(p_b))p_b, \qquad \text{and} \qquad 1-F(\overline{v})=(k_g+h_g(p_g))p_g,$$

whereas if $\underline{v}=\overline{v}$, the cut-off is given by v^* satisfying (1)-(2) as in the baseline setting. The cut-offs are unique under the following assumption which refines assumption (A_1).

Assumption (A_1')**.** For all $\overline{v}$ and $\underline{v}$ such that $\overline{v}\geq\underline{v}$,

$$\begin{cases} 1+\mu\dfrac{\partial}{\partial\overline{v}}\Big(\mathcal{M}^+(\overline{v})-\mathbb{E}[v|\underline{v}\leq v<\overline{v}]\Big)\geq 0, \\ 1+\mu\dfrac{\partial}{\partial\underline{v}}\Big(\mathbb{E}[v|\underline{v}\leq v<\overline{v}]-\mathcal{M}^-(\underline{v})\Big)\geq 0 \end{cases}$$

By constrast, if $[k_b+k_g+h_g(c_b)]c_b>1$ (and thus $p_b<c_b$), an equilibrium is described by a cut-off $\tilde{v}$ such that types $v\geq\tilde{v}$ invest in the brown technology and clean it, while types $v<\tilde{v}$ invest indifferently in the green or in the brown technology (leaving the latter dirty). The cut-off $\tilde{v}$ is given by the investors' segmentation condition:

$$\frac{(e-\tilde{e})}{p_b}[\tilde{v}+\mu\Delta(\tilde{v})]-\frac{\delta}{p_b}=0,$$

together with the price parity condition $p_g=p_b$ $(<c_b)$. Hence, assuming (A_1), the cut-off $\tilde{v}$ is unique.

Proposition B.6. ***(Investing for change)*** *The pecking order induced by direct consequentialism is given by:*

(i) If $[k_b+k_g+h_g(c_b)]c_b\leq 1$, then: 1^o investing in the green technology (highest types), 2^o investing in the brown technology and cleaning it (intermediate types),3^o investing in brown technology and not cleaning it (lowest types);

[59] The expressions for the perceived induced externalities stem from the fact that from a direct consequentialist perspective, investors should consider the next most attractive investment option as a counterfactual when assessing the induced externality of their saving decision. Here, each investor in technology $s\in\{b,g\}$ owns a firm producing $1/p_s$ unit of the consumption good. An investor with type $\underline{v}$ is indifferent between cleaning or not her brown firm: since her firm produces $1/p_b$ units of the consumption good, the total externality generated by the firm is equal to $\tilde{e}/p_b$ if cleaned, and e/p_b otherwise.

(ii) *By contrast, if* $[k_b + k_g + h_g(c_b)]c_b > 1$*, then:* 1^o *investing in the brown technology and cleaning it (highest types),* 2^o *(ex aequo) investing in the green technology or investing in brown technology (intermediate & lower types).*

The same intuition as the one underlying Proposition 4 applies.

B.4 Proof of Lemma B.1

Let $p_b \geq c_b$ and $p_g \geq c_g$. An equilibrium (v^*, p_b, p_g) is given by the following system of equations:

$$\begin{cases} e^\dagger [v^* + \mu\Delta(v^*)] - \dfrac{1}{p_b} + \dfrac{1}{p_g} = 0, \\ F(v^*) - (k_b + h_b(p_b))p_b = 0, \\ 1 - F(v^*) - (k_g + h_g(p_g))p_g = 0. \end{cases}$$

The associated Jacobian matrix writes as

$$J \equiv \begin{pmatrix} e^\dagger [1 + \mu\Delta'(v^*)] & \dfrac{de^\dagger}{dp_b}[v^* + \mu\Delta(v^*)] + \dfrac{1}{p_b^2} & -\dfrac{1}{p_g^2} \\ f(v^*) & -[k_b + h_b(p_b) + p_b h_b'(p_b)] & 0 \\ -f(v^*) & 0 & -[k_g + h_g(p_g) + p_g h_g'(p_g)] \end{pmatrix}$$

Its determinant is equal to

$$|J| = f(v^*)[k_b + h_b(p_b) + p_b h_b'(p_b)][k_g + h_g(p_g) + p_g h_g'(p_g)]D,$$

with

$$\begin{aligned} D &= e^\dagger[1 + \mu\Delta'(v^*)]\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{f(v^*)} + \frac{de^\dagger}{dp_b}[v^* + \mu\Delta(v^*)] + \frac{1}{p_b^2} + \frac{1}{p_g^2}\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{k_g + h_g(p_g) + p_g h_g'(p_g)} \\ &= e^\dagger[1 + \mu\Delta'(v^*)]\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{f(v^*)} + \frac{1}{e^\dagger}\frac{de^\dagger}{dp_b}\left(\frac{1}{p_b} - \frac{1}{p_g}\right) + \frac{1}{p_b^2} + \frac{1}{p_g^2}\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{k_g + h_g(p_g) + p_g h_g'(p_g)} \end{aligned}$$

where the second equality derives from the investors' segmentation condition (1). Hence J is invertible if and only if $D \neq 0$.

With the main specification of the perception of the induced externality (3),

$$\frac{1}{e^\dagger}\frac{de^\dagger}{dp_b}\left(\frac{1}{p_b} - \frac{1}{p_g}\right) = -\frac{1}{p_b}\left(\frac{1}{p_b} - \frac{1}{p_g}\right),$$

Therefore, $D > 0$, which yields the result.

By contrast, other situations may arise if the perceived induced externality is instead given by (4). If $de^\dagger/dp_b \geq 0$, then assumption (A_1) implies that $D > 0$. By contrast, if $de^\dagger/dp_b < 0$, then assuming (A_1), a sufficient condition for $D > 0$ is that

$$\frac{1}{e^\dagger}\frac{de^\dagger}{dp_b}\frac{1}{p_b} + \frac{1}{p_b^2} > 0, \qquad \text{i.e.} \qquad \frac{d}{dp_b}(p_b e^\dagger) > 0$$

Assumption (A_2) then provides a necessary and sufficient condition for the above inequality to hold. Indeed, by totally differentiating the market clearing conditions (2),

$$f(v^*)dv^* = [k_b + h_b(p_b) + p_b h_b'(p_b)]dp_b = -[k_g + h_g(p_g) + p_g h_g'(p_g)]dp_g, \tag{9}$$

and thus by definition of the perceived induced externality,

$$p_b e^\dagger = \frac{ep_b di_b}{f(v^*)dv^*} = \frac{ep_b h_b'(p_b)}{k_b + h_b(p_b) + p_b h_b'(p_b)}$$

As a consequence, $p_b e^\dagger$ increases with $p_b \geq c_b$ if and only if

$$\frac{p_b h_b'(p_b)}{k_b + h_b(p_b)} \qquad \text{increases with } p_b \geq c_b,$$

which is what assumption (A_2) requires.

Therefore, under assumptions (A_1) and (A_2), the implicit function theorem applies, and there exists a unique cut-off v^* and unique equilibrium prices p_b and p_g solving (1)-(2).

C Consumers with ethical preferences

We assumed so far that the consumption good had the same value to investors whether produced with a clean or dirty technology. Put differently, we assumed that consumers had no ethical concerns. Such a view is challenged by substantial evidence. Would our insights change if consumers were ethically motivated and valued differently the consumption good depending on its production technology?

Building on the baseline model, we assume there is a continuum of ethically-motivated consumers with intrinsic type ν drawn from a full-support distribution F_c on $\mathbb{R}_+$. A type-ν consumer derives an intrinsic motivation for consuming a good produced with the green technology equal to $e_c\nu$ with $e_c > 0$ denoting the consumers' environmental sensitivity (the case $e_c = 0$, i.e. non-ethically motivated consumers, corresponds to our former analysis). We assume for the sake of simplicity that consumers have no image concerns. Investors derive no utility from the consumption good *per se*, and thus sell all units they own to consumers. Each consumer has a fixed endowment of 1 unit of a numéraire, which she spends on green or brown consumption goods. In equilibrium, there exists a cut-off type ν^* below which all consumers buy the brown consumption good ("brown consumers"), and above which they buy the green consumption good ("green consumers"). We assume that sellers of consumption goods cannot price-discriminate consumers. Letting $\tilde{p}_b^c$ and $\tilde{p}_g^c$ be the prices of brown and green consumption goods, market clearing implies that a consumer with type $\nu < \nu^*$ (resp. $\nu \geq \nu^*$) buys $1/\tilde{p}_b^c = (k_b + i_b)/F(\nu^*)$ units of consumption good (resp. $1/\tilde{p}_g^c = (k_g + i_g)/(1 - F(\nu^*))$ units of consumption good). Hence the consumers' segmentation is determined by

$$e_c \nu^* - \frac{1}{\tilde{p}_b^c} + \frac{1}{\tilde{p}_g^c} = 0, \tag{10}$$

and thus "green consumers" forego a higher level of consumption with respect to "brown consumers" – we refer to the (stricly positive) consumption differential between brown and green consumers as the *consumers' green premium.* The investors' segmentation is thus given by

$$e^\dagger \left[v^* + \mu \Delta(v^*) \right] - \frac{\tilde{p}_b^c}{p_b} + \frac{\tilde{p}_g^c}{p_g} = 0,$$

i.e. equivalently, using the market clearing conditions (2),

$$e^\dagger \left[v^* + \mu \Delta(v^*) \right] - \frac{F_c(\nu^*)}{F(v^*)} + \frac{1 - F_c(\nu^*)}{1 - F(v^*)} = 0 \tag{11}$$

The consumers' and investors' segmentation conditions (10)-(11), together with the market clearing equations (2) and the expression for the induced externality (either (3) (or (4))) characterize a candidate equilibrium. Assuming (A_1), this equilibrium is uniquely defined under the following (sufficient) assumption, which may be interpreted as requiring the induced externality not to vary too steeply with respect to the brown investment level:

Assumption (A_3)**.** $de^\dagger/di_b \in (-e, e)$ for any $p_b \geq c_b$, where $e^\dagger$ is given by (3) (or (4)).

Ethically-motivated consumers lower the investors' green premium because of their readiness to consume green goods rather than brown ones, yet the investors' green premium remains positive. Under (A_1) and (A_3), the same comparative statics as in the baseline model obtain (see Proposition 1): a higher environmental sensitivity either on the consumers' or the investors' side yields more "green" behaviour on both sides (more green investments and more green consumption). The consumers' green premium increases with their (or investors') environmental sensitivity, whereas the comparative statics of the investors' green premium are ambiguous. As intuitive, upward shifts in the consumers' environmental sensitivity (modelled for instance as a uniform shift in their type distribution) lowers the investors' green premium.

Taxation and ethically-motivated consumers. For the sake of comparison, we assume the principal attributes the same value to one unit of the consumption good irrespectively of its production technology. Its welfare objective is thus still given by W, yielding whenever $C_b'(0) \leq 1-e$, the first-best prices $p_g = 1$ and $p_b = 1-e$. Suppose the principal considers taxing the investors' financial returns, and that the latter internalize their impact on tax proceeds (broad internalization). Then, in contrast to the results in Proposition 2, the Pigouvian tax does not achieve the first-best investments,[60] as consumers now provide an incentive for investors to save in the green technology.

Lastly, ethically-motivated consumers alter the political economy of divestment strategies (as well as exclusion and best-in-class strategies), as the latter now bear differentiated consequences for consumers: public entities' divestment to the green technology makes green consumers better off, and brown consumers worse off.

D Decommissioning

We extend the baseline setting by allowing production operators to decommission existing capital stocks. We assume that (i) operating installed capital stocks involves an operation/maintenance cost per unit of capital, and that (ii) decommissioning an installed capital unit is costly. The decommissioning cost may stem for instance from a regulation requiring the plant to be deconstruct, rather than simply abandoned, and the area to be cleaned in order to safeguard the environment. We assume that there exists a price level such that for any lower price, decommissioning one unit of capital stock is more profitable than operating it, whereas for any higher price, operating the capital unit is more profitable.

Hence, for technology $s \in \{b, g\}$, there exists a price level $c_s^d \leq c_s$ and a function $\tilde{h}_s$ of the price p_s defined over $[0, c_s^d]$ with $h(0) = 0$ and $\tilde{h}_s(c_s^d) = k_s$, such that the overall level of

[60] Indeed, whenever $C_b'(0) \leq 1-e$, the investors' segmentation condition at the first-best investment prices writes as

$$\left(1-\frac{\tau}{e}\right)e^{\dagger}\left[v^* + \mu\Delta(v^*)\right] - \frac{1-\tau}{1-e}\left[\frac{F_c(\nu^*)}{k_b + h_b(1-e)}\right] + \frac{1-F_c(\nu^*)}{k_g + h_g(1)} = 0$$

investment in technology s writes as the following function of the price p_s:

$$h_s^d(p_s) = \begin{cases} \tilde{h}_s(p_s) & \text{if } p_s \leq c_s^d, \\ k_s & \text{if } c_s^d < p_s < c_s, \\ k_s + h_s(p_s) & \text{if } p_s \geq c_s \end{cases}$$

We assume that $\tilde{h}_s$ is strictly increasing and twice continuously differentiable.

Consistently with direct consequentialism, with the main specification of the perceived induced externality (3), investors now take the induced externality of saving in the brown rather than in the green technology as equal to

$$e^\dagger = \frac{e}{p_b}\left(1 - \mathbf{1}_{\{p_b \in (c_s^d, c_s)\}}\right)$$

Similarly, the alternative specification for the perceived induced externality (4) now writes as

$$e^\dagger = \frac{e dh_b^d(p_b)}{f(v^*)dv^*},$$

and thus, using the brown market clearing condition,

$$e^\dagger = \begin{cases} \dfrac{e\tilde{h}_b(p_b)}{\tilde{h}_b(p_b) + p_b\tilde{h}_b(p_b)} & \text{if } p_b \leq c_b^d, \\ 0 & \text{if } c_b^d < p_b < c_b, \\ \dfrac{eh_b'(p_b)}{k_b + h_b(p_b) + p_b h_b'(p_b)} & \text{if } p_b \geq c_b \end{cases}$$

In contrast with the baseline setting, $e^\dagger > 0$ whenever $p_b \leq c_b^d$: impact investors take into account that the brown capital units they may buy would be otherwise decommissioned. Hence, from a direct consequentialist perspective, they are responsible for the pollution generated by those units. Nonetheless, $e^\dagger$ remains nil on the range of prices (c_b^d, c_b) such that brown capital units would not be decommissioned anyway. The results noted in the text, thus derive from the same arguments as the ones used when production operators cannot decommission existing capital stocks – which may now be interpreted as the case $c_s^d = 0$. In particular, *neutrality results* now only obtain whenever the brown price p_b is in the range (c_b^d, c_b), while the usual mechanisms apply whenever $p_b \leq c_b^d$ or $p_b \geq c_b$.

E Proof of Lemma 1

The result follows from the investors' segmentation condition (1), and in the case where the perceived induced externality is given by (4), from the assumption that $h'_b(p_b) > 0$ for $p_b \in [c_b, +\infty)$. Hence, the green premium is strictly positive if and only if $e^\dagger > 0$, i.e. if and only if $p_b \geq c_b$. If $p_b < c_b$, then the green premium is nil and thus $p_g = p_b = p < c_b$. Hence, since $c_b < c_g$, there is no investment in either technology ($i_b = i_g = 0$), and the equilibrium price is given by the market clearing condition: $p\big[k_b + k_g + h_g(p)\big] = 1$.

F Proofs of Propositions 1 and A.1

With the naive specification of the induced externality (3), Proposition 1 follows from the following, more complete Lemma, which we write in a stylized form.

Lemma F.1 (Comparative statics for the equilibrium prices and green premium).

- *With respect to the (unit) externality e: if $e \uparrow$, then $p_b \downarrow$, $1 - F(v^*) \uparrow$, $p_g \uparrow$, $e^\dagger \uparrow$, $\gamma \uparrow$*
- *With respect to green capital stock: if $k_g \uparrow$, then $p_b \downarrow$, $1 - F(v^*) \uparrow$, $p_g \downarrow$, $e^\dagger \uparrow$*
- *With respect to brown capital stock: if $k_b \uparrow$, then $p_b \downarrow$, $1 - F(v^*) \downarrow$, $p_g \downarrow$, $e^\dagger \uparrow$, $\gamma \uparrow$*
- *With respect to the green (resp. brown) investment supply function:if $h_g \uparrow$ (resp. if $h_b \uparrow$),[61] then $p_b \downarrow$, $1 - F(v^*) \uparrow$, $p_g \downarrow$, $e^\dagger \uparrow$, and $i_g \uparrow$ (resp. $i_b \uparrow$ and $\gamma \uparrow$)*
- *With respect to the brown investment supply function: if $h_b \uparrow$, then $p_b \downarrow$, $e^\dagger \uparrow$, $1-F(v^*) \downarrow$, $p_g \downarrow$, $i_b \uparrow$*
- *With respect to the distribution of types (willingness to do good): if $F \uparrow$ in terms of first-order stochastic dominance, then $p_b \downarrow$, $p_g \uparrow$, $e^\dagger \uparrow$, $\gamma \uparrow$.*

Let us begin with the proof of the comparative static with respect to the distribution of types F, mentioned in the exposition. Let F and G denote two c.d.f. on $\mathbb{R}_+$ such that G strictly dominates F in the first-order stochastic sense, i.e. such that for all $v \in \mathbb{R}_+^*$, $G(v) < F(v)$. Suppose by contradiction that the respective cutoffs v_G^* and v_F^* are such that $F(v_F^*) \leq G(v_G^*)$. Then, by market clearing, $p_{b,F} \leq p_{b,G}$ and $p_{g,F} \geq p_{g,G}$. However, $F(v_F^*) \leq G(v_G^*)$ implies by

[61] We consider homothetic variations in the technologies costs: namely, we write the investment functions as $\alpha_g h_g$ and $\alpha_b h_b$ with α_g and α_b two strictly positive constants, and explore the comparative statics with respect to α_g and α_b. Changes in the cost functions may be due for instance to a technological shock or a change in the regulation.

strict FOSD, $v_F^* < v_G^*$. Hence,

$$ev_F^* < ev_G^* \qquad \text{and} \qquad 1 - \frac{p_{b,F}}{p_{g,F}} \geq 1 - \frac{p_{b,G}}{p_{g,G}}$$

This contradicts the equilibrium condition (1):

$$ev_F^* - \left(1 - \frac{p_{b,F}}{p_{g,F}}\right) = ev_G^* - \left(1 - \frac{p_{b,G}}{p_{g,G}}\right) = 0.$$

Therefore, $F(v_F^*) > G(v_G^*)$ and thus $p_{b,F} > p_{b,G}$ and $\gamma_F < \gamma_G$. (Note that with the soophisticated perception of the induced externality, $e^\dagger_{\text{soph}}$ given by (4), Assumption (A_2) ensures that the same argument applies, as then $p_b e^\dagger_{\text{soph}}$ increases with p_b.)

The proof for the other comparative statics (in the general case of image concerns and with naive or sophisticated agents) relies on (9) and total differentiation of the LHS of the investors' segmentation condition (1). We first explicit the computations for the comparative statics with respect to the externality e when agents are sophisticated. The case $p_b < c_b$ is straightforward (as $e^\dagger = 0$). Consider the case $p_b \geq c_b$. Totally differentiating (1), and writing the terms dv^* and dp_g as functions of dp_b via the differentiated market clearing conditions (9) yields

$$\left[e^\dagger[1 + \mu\Delta'(v^*)]\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{f(v^*)} + \frac{1}{e^\dagger}\frac{\partial e^\dagger}{\partial p_b}\left(\frac{1}{p_b} - \frac{1}{p_g}\right) + \frac{1}{p_b^2} + \frac{1}{p_g^2}\frac{k_b + h_b(p_b) + p_b h_b'(p_b)}{k_g + h_g(p_g) + p_g h_g'(p_g)}\right]dp_b$$
$$= -\frac{h_b'(p_b)}{k_b + h_b(p_b) + p_b h_b'(p_b)}\left[v^* + \mu\Delta(v^*)\right]de$$

Assumptions (A_1) and (A_2) thus imply that $dp_b/de < 0$. By the market clearing conditions and their total differentiation (9), this further implies that $dv^*/de < 0$, $dp_g/de > 0$, and thus $d\gamma/de > 0$. Lastly, by totally differentiating (3) (or (4)),

$$de^\dagger = \frac{\partial e^\dagger}{\partial p_b}dp_b + \frac{e^\dagger}{e}de$$

Assuming (A_1) (and if the perceived induced externality is given by (4) instead of (3), assuming

(A_2)), and using the above expression for dp_b/de gives that $de^\dagger/de$ has the same sign as

$$
\begin{aligned}
&-\frac{h_b'(p_b)}{k_b+h_b(p_b)+p_bh_b'(p_b)}\left[v^*+\mu\Delta(v^*)\right]\frac{\partial e^\dagger}{\partial p_b}\\
&+\frac{e^\dagger}{e}\left[e^\dagger[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1}{e^\dagger}\frac{\partial e^\dagger}{\partial p_b}\left(\frac{1}{p_b}-\frac{1}{p_g}\right)+\frac{1}{p_b^2}+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\right]\\
&=\frac{e^\dagger}{e}\left[e^\dagger[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1}{p_b^2}+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\right]
\end{aligned}
$$

where the equality derives from the investors' segmentation condition (1). Hence, $de^\dagger/de>0$.

The other comparative statics obtain similarly. As an illustration, we provide some further details in the case where the induced externality is given by (3). Consider the comparative static with respect to k_b. Following the same steps as previously yields

$$
\begin{aligned}
&\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}+\frac{1}{p_bp_g\left[k_b+h_b(p_b)+p_bh_b'(p_b)\right]}+\frac{1}{p_g^2\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}\right]dp_b\\
&\quad=-\frac{p_b}{k_b+h_b(p_b)+p_bh_b'(p_b)}\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}+\frac{1}{p_g^2\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}\right]dk_b.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}+\frac{1}{p_bp_g\left[k_b+h_b(p_b)+p_bh_b'(p_b)\right]}+\frac{1}{p_g^2\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}\right]f(v^*)dv^*\\
&=-\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}+\frac{1}{p_bp_g\left[k_b+h_b(p_b)+p_bh_b'(p_b)\right]}+\frac{1}{p_g^2\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}\right]\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]dp_g\\
&=\frac{1}{p_g\left[k_b+h_b(p_b)+p_bh_b'(p_b)\right]}dk_b,
\end{aligned}
$$

and the comparative statics of the green premium obtain by the investors' segmentation condition. [By contrast, if the perceived induced externality is given by (4), these comparative statics become ambiguous as $e^\dagger$ depends explicitly on k_b.]

Similarly, considering the comparative statics with respect to k_g and following the same steps yields

$$
\begin{aligned}
&\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}\left[k_b+h_b(p_b)+p_bh_b'(p_b)\right]+\frac{1}{p_bp_g}+\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{p_g^2\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}\right]dp_b\\
&\quad=-\frac{1}{p_g\left[k_g+h_g(p_g)+p_gh_g'(p_g)\right]}dk_g.
\end{aligned}
$$

Hence,

$$\begin{aligned} dp_g &= p_g^2\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}[k_b+h_b(p_b)+p_bh_b'(p_b)]+\frac{1}{p_bp_g}\right]dp_b \\ &= p_g^2\left[e^\dagger\frac{1+\mu\Delta'(v^*)}{f(v^*)}[k_b+h_b(p_b)+p_bh_b'(p_b)]+\frac{1}{p_bp_g}\right]\frac{f(v^*)}{k_b+h_b(p_b)+p_bh_b'(p_b)}dv^*. \end{aligned}$$

Lastly, as noted in the text, the comparative statics of the aggregate level of pollution with respect to the unit externality e are unclear in general:

$$\frac{d}{de}[e(k_b+h_b(p_b))] = k_b+h_b(p_b)+eh_b'(p_b)\frac{dp_b}{de} \tag{12}$$

Hence, assuming (A_1) (and if the perceived induced externality is given by (4) instead of (3), assuming (A_2)), and using the investors' segmentation condition (1), the LHS has the same sign as

$$\begin{aligned} &\left[e^\dagger[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1}{e^\dagger}\frac{\partial e^\dagger}{\partial p_b}\left(\frac{1}{p_b}-\frac{1}{p_g}\right)+\frac{1}{p_b^2}+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\right] \\ &\quad -\frac{h_b'(p_b)}{k_b+h_b(p_b)}\left(\frac{1}{p_b}-\frac{1}{p_g}\right) \end{aligned}$$

The same condition holds for the two moral criteria introduced in Section 4, replacing $e^\dagger$ with $e^\dagger_{rc}$ (which is a constant) or $e^\dagger_{sr} = e/p_b$. Hence the same ambiguity prevails with these two moral criteria. In the case of rule consequentialism, a sufficient condition for a lower unit externality to induce a lower aggregate level of pollution is that $p_b/[k_b+h_b(p_b)]$ be increasing with p_b (which may be reminiscent of assumption (A_2)).

Nevertheless, for all three moral criteria, (12) implies that for e sufficiently small (close to zero), a lower unit externality generates a lower aggregate level of pollution (i.e. a lower externality level).

G Proofs of Section 3.1

G.1 Proof of comparative statics with taxes

The result derives from computations analogous to the ones of Appendix F. With narrow internalization, totally differentiating (8) yields

$$\Big[e^\dagger[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1}{e^\dagger}\frac{\partial e^\dagger}{\partial p_b}\Big(\frac{1-\tau_b}{p_b}-\frac{1}{p_g}\Big)$$
$$+\frac{1-\tau_b}{p_b^2}+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\Big]dp_b=-\frac{d\tau_b}{p_b}$$

Hence, assumption (A_1) (together with (A_2) if the perceived induced externality is given by (4) instead of (3)) implies that $dp_b/d\tau_b<0$, and thus by totally differentiating the market clearing conditions and the expression of the induced externality, $dv^*/d\tau_b<0$, $dp_g/d\tau_b>0$ and thus $d\gamma/d\tau_b>0$. Notwithstanding, the comparative statics of the tax-adjusted green premium $(1-\tau_b)/p_b-1/p_g$ with respect to the tax rate are ambiguous in general. Indeed, totally differentating the tax-adjusted green premium yields

$$\frac{d}{d\tau_b}\Big(\frac{1-\tau_b}{p_b}-\frac{1}{p_g}\Big)=-\frac{1}{p_b}-\Big(\frac{1-\tau_b}{p_b^2}+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\Big)\frac{dp_b}{d\tau_b}$$
$$=-\frac{1}{p_b}\Big[e^\dagger[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1}{e^\dagger}\frac{\partial e^\dagger}{\partial p_b}\Big(\frac{1-\tau_b}{p_b}-\frac{1}{p_g}\Big)\Big],$$

and while assumption (A_1) implies that the first term between brackets on the RHS is positive, the second term is negative with naive investors (and can be negative with sophisticated investors despite assumption (A_2)), and thus the sign of the RHS is ambiguous in general. Nevertheless, whenever the tax-adjusted green premium is sufficiently small (close to 0), then it decreases with the tax rate. [Furthermore, for the two moral criteria introduced in Section 4, assuming only (A_1) in the presence of image concerns, we have that $\partial e^\dagger_{sr}/\partial p_b<0$ and $\partial e^\dagger_{rc}/\partial p_b=0$, and therefore this ambiguity persists with shared responsibility, yet disappears with rule consequentialism (the tax-adjusted green premium unambiguously decreases with the tax rate with rule consequentialist investors).]

With broad internalization, totally differentiating the investors' segmentation condition yields

$$\Big(\frac{e^\dagger}{e}[v^*+\mu\Delta(v^*)]-\frac{1}{p_b}\Big)\frac{d\tau_b}{dp_b}=\Big[(e^\dagger-t^\dagger)[1+\mu\Delta'(v^*)]\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{f(v^*)}+\frac{1-\tau_b}{p_b^2}$$
$$+\frac{1}{e^\dagger}\Big(1-\frac{\tau_b}{e}\Big)\frac{\partial e^\dagger}{\partial p_b}\Big(\frac{1-\tau_b}{p_b}-\frac{1}{p_g}\Big)+\frac{1}{p_g^2}\frac{k_b+h_b(p_b)+p_bh_b'(p_b)}{k_g+h_g(p_g)+p_gh_g'(p_g)}\Big]$$

For any $\tau_b < e$, assumption (A_1) (together with (A_2) if the perceived induced externality is given by (4) instead of (3)) implies that the RHS is strictly positive. As for the the left hand side, the investors' segmentation condition implies that:

$$\left(\frac{e^\dagger}{e}[v^* + \mu\Delta(v^*)] - \frac{1}{p_b}\right) = \frac{1}{e-\tau_b}\left(\frac{1-\tau_b}{p_b} - \frac{1}{p_g}\right) - \frac{1}{p_b} = \frac{1}{e-\tau_b}\left(\frac{1-e}{p_b} - \frac{1}{p_g}\right).$$

G.2 Proof of Propositions 2, A.2 and B.4, and Lemma 2

Let us first consider naive investors. With narrow internalization, the general optimal taxation problem writes as

$$\max_{\tau}\ (1-e)i_b - C_b(i_b) + i_g - C_g(i_g)$$

subject to

$$\begin{cases} p_b[k_b+i_b] + p_g[k_g+i_g] = 1, & \text{for all } p_b \geq 0, \\ \dfrac{e}{p_b}F^{-1}\Big(p_b[k_b+i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b \geq c_b, \\ \dfrac{1-\tau}{p_b} = \dfrac{1}{p_g} & \text{if } p_b < c_b. \end{cases}$$

Similarly, with broad internalization, the general optimal taxation problem writes as

$$\max_{\tau}\ (1-e)i_b - C_b(i_b) + i_g - C_g(i_g)$$

subject to

$$\begin{cases} p_b[k_b+i_b] + p_g[k_g+i_g] = 1, & \text{for all } p_b \geq 0, \\ \dfrac{e-\tau}{p_b}F^{-1}\Big(p_b[k_b+i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b \geq c_b \text{ and } \tau \leq e, \\ \dfrac{e-\tau}{p_b}F^{-1}\Big(1-p_b[k_b+i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b \geq c_b \text{ and } \tau > e, \\ -\dfrac{\tau}{p_b}F^{-1}\Big(1-p_b[k_b+i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b < c_b. \end{cases}$$

Consider the case $1-e \geq c_b$. Then, with narrow internalization, a necessary condition for a tax τ to implement the first-best prices is that investors' segmentation condition holds, i.e. that

$$e - \tau = eF^{-1}\Big(p_b[k_b+i_b]\Big) > 0,$$

while with broad internalization, the investors' segmentation condition at the first-best prices holds for the Pigouvian tax $\tau = e$. (By assumption, markets clear at first-best prices.) Similarly, consider the case $1 - e < c_b$. Then, with narrow internalization, the investors' segmentation condition at the first-best prices holds for the Pigouvian tax $\tau = e$, while with broad internalization, a necessary condition for a tax τ to implement the first-best prices is that

$$e - \tau = -\tau F^{-1}\Big(1 - p_b[k_b + i_b]\Big) < 0.$$

The proof for the result with the sophisticated expression of the perceived induced externality (4) follows from the same arguments as with the naive expression (3), using that for any brown price p_b, $t^{\dagger}_{\text{soph}} = (\tau_b/e)e^{\dagger}_{\text{soph}}$. Hence, with broad internalization, the general optimal taxation problem writes as

$$\max_{\tau}\ (1-e)i_b - C_b(i_b) + i_g - C_g(i_g)$$

subject to

$$\begin{cases} p_b[k_b + i_b] + p_g[k_g + i_g] = 1, & \text{for all } p_b \geq 0,\\ (e^{\dagger}_{\text{soph}} - t^{\dagger}_{\text{soph}})F^{-1}\Big(p_b[k_b + i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b \geq c_b \text{ and } \tau \leq e,\\ (e^{\dagger}_{\text{soph}} - t^{\dagger}_{\text{soph}})F^{-1}\Big(1 - p_b[k_b + i_b]\Big) = \dfrac{1-\tau}{p_b} - \dfrac{1}{p_g} & \text{if } p_b \geq c_b \text{ and } \tau > e,\\ \dfrac{1-\tau}{p_b} = \dfrac{1}{p_g} & \text{if } p_b < c_b. \end{cases}$$

Consequently, with broad internalization, the investors' segmentation condition at the first-best prices holds for the Pigouvian tax $\tau = e$ (and markets clear by assumption).

The same arguments apply with image concerns, assuming (A_1).

H Proof of Propositions 3 and B.5

We focus on the case in which parameter values e, μ and G such that $v^*_{pub} > 0$ (as otherwise there can be no public divestment).

Consider the setting in which $p^{\text{nd}}_b > c_b$, i.e. $i^{\text{nd}}_b > 0$. In order to show claim (i), we show that $i^{\text{d}}_b < i^{\text{nd}}_b$, i.e. equivalently that $p^{\text{d}}_b < p^{\text{nd}}_b$. Suppose by contradiction that $p^{\text{d}}_b \geq p^{\text{nd}}_b$. Then the market clearing conditions imply that $v^*_{priv} < v^*$ [62], and that $p^{\mathrm{d}}_g \leq p^{\mathrm{nd}}_g$. Hence, by the investors' segmentation conditions, we have that:

$$e\big[v^*_{priv} + \mu\Delta_F(v^*_{priv})\big] = 1 - \frac{p^{\mathrm{nd}}_b}{p^{\mathrm{nd}}_g} \geq 1 - \frac{p^{\mathrm{d}}_b}{p^{\mathrm{d}}_g} = e\big[v^* + \mu\Delta_F(v^*)\big]$$

Similarly, if the perceived induced externality is given by (4) instead of (3), then the investors' segmentation conditions together with (A_2) imply that:[63]

$$v^*_{priv} + \mu\Delta_F(v^*_{priv}) \geq v^* + \mu\Delta_F(v^*)$$

Hence assumption (A_1) implies that $v^*_{priv} \geq v^*$, which is a contradiction.

Consequently, $p^{\mathrm{d}}_b < p^{\mathrm{nd}}_b$ and thus $i^{\mathrm{d}}_b < i^{\mathrm{nd}}_b$, i.e. divestment decreases the aggregate externality level. Moreover, as $p^{\mathrm{d}}_b < p^{\mathrm{nd}}_b$, divestment raises the brown financial returns and lowers the green financial returns. Lastly, by the same arguments as above, as $p^{\mathrm{d}}_g > p^{\mathrm{nd}}_g$, the investors' segmentation conditions (together with assumption (A_2) if the perceived induced externality is given by (4)) yield that:

$$v^*_{priv} + \mu\Delta_F(v^*_{priv}) < v^* + \mu\Delta_F(v^*),$$

and thus assumption (A_1) implies that $v^*_{priv} < v^*$.

By contrast, consider the setting in which $p^{\mathrm{nd}}_b \leq c_b$, i.e. $i^{\mathrm{nd}}_b = 0$. Proceeding by contradiction and using the same arguments as above yields that $p^{\mathrm{d}}_b \leq p^{\mathrm{nd}}_b$. Therefore, $i^{\mathrm{d}}_b = 0 = i^{\mathrm{nd}}_b$.

Changing the investment mandate of public entities. Assume that the government sets the distribution of public types as $G(v) = F(v - \theta)$ with the parameter $\theta > 0$ governing how green public entities should be. Comparative statics of the equilibrium variables with respect to θ obtain by the usual computations, whenever μ is sufficiently low for the variations of $\mu\Delta$ with

[62] Indeed, by assumption, $\alpha G(v^*_{pub}) > 0$, while by the brown market clearing conditions, we get that:

$$\alpha G(v^*_{pub}) + (1-\alpha)F(v^*_{priv}) = (k_b + h_b(p^{\mathrm{nd}}_b))p^{\mathrm{nd}}_b \leq (k_b + h_b(p^{\mathrm{d}}_b))p^{\mathrm{d}}_b = (1-\alpha)F(v^*).$$

[63] Namely, assumption (A_2) gives the following inequality:

$$\begin{aligned} e\big[v^*_{priv} + \mu\Delta_F(v^*_{priv})\big] &= \frac{k_b + h_b(p^{\mathrm{nd}}_b) + p^{\mathrm{nd}}_b h'_b(p^{\mathrm{nd}}_b)}{p^{\mathrm{nd}}_b h'_b(p^{\mathrm{nd}}_b)}\left(1 - \frac{p^{\mathrm{nd}}_b}{p^{\mathrm{nd}}_g}\right) \\ &\geq \frac{k_b + h_b(p^{\mathrm{d}}_b) + p^{\mathrm{d}}_b h'_b(p^{\mathrm{d}}_b)}{p^{\mathrm{d}}_b h'_b(p^{\mathrm{d}}_b)}\left(1 - \frac{p^{\mathrm{d}}_b}{p^{\mathrm{d}}_g}\right) = e\big[v^* + \mu\Delta_F(v^*)\big]. \end{aligned}$$

respect to θ not to offset the downward impact of θ on brown investment. Whenever it is so, a higher θ (i.e. a greener investment mandate) induces a lower aggregate level of brown investment and thus a lower externality level.

I Proof of Propositions 4 and B.6

The perceived induced externality of investing in the brown or cleaned brown technologies rather than in the green one hinges on the investors' degree of sophistication (see footnote 64 below). By contrast, the induced externality of investing in the standard brown technology instead of the cleaned brown one is given by $(e - \tilde{e})/p_b$, independently of the investors' degree of sophistication. Indeed, the brown price is fixed when the decision to clean or not brown shares is taken (step 2), which neutralizes indirect equilibrium effect on prices. Each brown investor owns $1/p_b$ units of brown investment. An investor with type $\underline{v}$ is indifferent between cleaning or not her brown firm: since her firm produces $1/p_b$ units of the consumption good, the total externality generated by the firm is equal to $\tilde{e}/p_b$ if cleaned, and e/p_b otherwise.

Hence an equilibrium is described by two cut-offs such that $\overline{v} \geq \underline{v}$ such that types above $\overline{v}$ invest in the green technology, types $v \in [\underline{v}, \overline{v}]$ invest in the brown technology and clean it, and types below $\underline{v}$ invest in the brown technology and keep it dirty. If $\underline{v} < \overline{v}$, the investors' segmentation conditions write as[64]

$$\begin{cases} \dfrac{\tilde{e}}{p_b}\left[\overline{v} + \mu\big(\mathcal{M}^+(\overline{v}) - \mathbb{E}[v|\underline{v} \leq v < \overline{v}]\big)\right] - \left(\dfrac{1-\delta}{p_b} - \dfrac{1}{p_g}\right) = 0 \\ \dfrac{(e-\tilde{e})}{p_b}\left[\underline{v} + \mu\big(\mathbb{E}[v|\underline{v} \leq v < \overline{v}] - \mathcal{M}^-(\underline{v})\big)\right] - \dfrac{\delta}{p_b} = 0, \end{cases}$$

together with the market clearing conditions:

$$F(\overline{v}) = (k_b + h_b(p_b))p_b, \qquad \text{and} \qquad 1 - F(\overline{v}) = (k_g + h_g(p_g))p_g,$$

[64] Similarly, if the perceived induced externality is given by (4) instead of (3),

$$\begin{cases} \dfrac{\tilde{e} di_b}{f(\overline{v})d\overline{v}}\left[\overline{v} + \mu\big(\mathcal{M}^+(\overline{v}) - \mathbb{E}[v|\underline{v} \leq v < \overline{v}]\big)\right] - \left(\dfrac{1-\delta}{p_b} - \dfrac{1}{p_g}\right) = 0 \\ \dfrac{(e-\tilde{e})}{p_b}\left[\underline{v} + \mu\big(\mathbb{E}[v|\underline{v} \leq v < \overline{v}] - \mathcal{M}^-(\underline{v})\big)\right] - \dfrac{\delta}{p_b} = 0, \end{cases}$$

whereas if $\underline{v} = \overline{v}$, the unique cut-off is given by v^* satisfying (1)-(2) as in the baseline setting.

Uniqueness. An equilibrium $(\overline{v}, \underline{v}, p_b, p_g)$ is a solution to the system of equations

$$\begin{cases} \tilde{e}^\dagger \left[\overline{v} + \mu\overline{\Delta}(\underline{v}, \overline{v})\right] - \left(\dfrac{1-\delta}{p_b} - \dfrac{1}{p_g}\right) = 0 \\ (e - \tilde{e})\left[\underline{v} + \mu\underline{\Delta}(\underline{v}, \overline{v})\right] - \delta = 0 \\ F(\overline{v}) - (k_b + h_b(p_b))p_b = 0 \\ 1 - F(\overline{v}) - (k_g + h_g(p_g))p_g = 0 \end{cases}$$

where

$$\tilde{e}^\dagger \equiv e h_b'(p_b) / \left[k_b + h_b(p_b) + p_b h_b'(p_b)\right]$$
$$\overline{\Delta}(\underline{v}, \overline{v}) \equiv \mathcal{M}^+(\overline{v}) - \mathbb{E}[v | \underline{v} \le v < \overline{v}]$$
$$\underline{\Delta}(\underline{v}, \overline{v}) = \mathbb{E}[v | \underline{v} \le v < \overline{v}] - \mathcal{M}^-(\underline{v})$$

The associated Jacobian matrix writes as

$$J \equiv \begin{pmatrix} \tilde{e}^\dagger \left[1 + \mu \dfrac{\partial \overline{\Delta}}{\partial \overline{v}}\right] & \tilde{e}^\dagger \mu \dfrac{\partial \overline{\Delta}}{\partial \underline{v}} & \dfrac{d\tilde{e}^\dagger}{dp_b}\left[\overline{v} + \mu\overline{\Delta}(\underline{v}, \overline{v})\right] + \dfrac{1-\delta}{p_b^2} & -\dfrac{1}{p_g^2} \\ (e - \tilde{e})\mu \dfrac{\partial \underline{\Delta}}{\partial \overline{v}} & (e - \tilde{e})\left[1 + \mu \dfrac{\partial \underline{\Delta}}{\partial \underline{v}}\right] & 0 & 0 \\ f(\overline{v}) & 0 & -\varphi_b(p_b) & 0 \\ -f(\overline{v}) & 0 & 0 & -\varphi_g(p_g) \end{pmatrix}$$

Hence its determinant is equal to

$$\begin{aligned} |J| = {} & \tilde{e}^\dagger \left[1 + \mu \frac{\partial \overline{\Delta}}{\partial \overline{v}}\right](e - \tilde{e})\left[1 + \mu \frac{\partial \underline{\Delta}}{\partial \underline{v}}\right]\varphi_b(p_b)\varphi_g(p_g) - (e - \tilde{e})\mu \frac{\partial \underline{\Delta}}{\partial \overline{v}} \tilde{e}^\dagger \mu \frac{\partial \overline{\Delta}}{\partial \underline{v}} \varphi_b(p_b)\varphi_g(p_g) \\ & + f(\overline{v})(e - \tilde{e})\left[1 + \mu \frac{\partial \underline{\Delta}}{\partial \underline{v}}\right]\left(\frac{d\tilde{e}^\dagger}{dp_b}\left[\overline{v} + \mu\overline{\Delta}(\underline{v}, \overline{v})\right] + \frac{1-\delta}{p_b^2}\right)\varphi_g(p_g) + f(\overline{v})(e - \tilde{e})\left[1 + \mu \frac{\partial \underline{\Delta}}{\partial \underline{v}}\right]\varphi_b(p_b)\frac{1}{p_g^2} \end{aligned}$$

Therefore, assumption (A_1') (together with (A_2) if the perceived induced externality is given by (4)) implies that $|J| > 0$ [65], and thus J is invertible. As a consequence, under (A_1') (and (A_2) if the perceived induced externality is given by (4)), the above equilibrium is unique.

Suppose now that $\left[k_b + k_g + h_g(c_b)\right]c_b > 1$, i.e. there are no additional brown investments

[65] Indeed, by construction,

$$\frac{\partial \overline{\Delta}}{\partial \underline{v}} \le 0, \qquad \text{and} \qquad \frac{\partial \underline{\Delta}}{\partial \overline{v}} \ge 0$$

on top of the existing capital stock k_b. Hence investing in the brown technology instead of the green technology does not affect the aggregate level of brown capital units. Therefore, investing in the green technology instead of the dirty brown technology has no impact on the externality level: the shares of brown technology will be bought and used for production anyway – whether their buyer cleans them or not cannot enter a direct consequentialist's impact evaluation. Hence, from a direct consequentialist perspective, as pollution from the k_b units of brown capital stock happens anyway, investing in the brown technology and cleaning it, instead of keeping it dirty or investing in the green technology, yields a downward impact on the externality level, reducing pollution.

As a consequence, there exists a cut-off type such that types above the cut-off invest in the brown technology and clean it, while types below the cut-off are indifferent between investing in the green technology or in the brown one, keeping it dirty. By the latter's indifference, in equilibrium $p_b = p_g < c_b$. As noted in the text, whenever interior, the cut-off $\tilde{v}$ is thus given by

$$\frac{(e-\tilde{e})}{p_b}\big[\tilde{v}+\mu\Delta(\tilde{v})\big] - \frac{\delta}{p_b} = 0, \qquad \text{i.e.} \qquad (e-\tilde{e})\big[\tilde{v}+\mu\Delta(\tilde{v})\big] = \delta$$

while the market clearing conditions write as

$$1 - F(\tilde{v}) + (1-x)F(\tilde{v}) = k_b p_b, \qquad \text{and} \qquad xF(\tilde{v}) = (k_g + h_g(p_g))p_g$$

for some $x \in [0,1]$. As a consequence, assumption (A_1) yields the cut-off's uniqueness.

J Proof of Proposition 5

The result derives from the observation that all previous conclusions in the case $[k_b + k_g + h_g(c_b)]c_b \leq 1$ hinged on the (sufficient) condition:

$$\frac{d(p_b e^{\dagger})}{dp_b} \geq 0,$$

being satisfied. This condition holds if the perceived induced externality is given by (3), and under assumption (A_2), if the perceived induced externality is given by (4).

This condition also clearly holds with shared responsibility (with perceived induced externality $e^\dagger_{sr}$) and rule consequentialism (with perceived induced externality $e^\dagger_{rc}$)[66]. Hence all the previous results hinging on $e^\dagger > 0$ still hold with rule consequentialism and shared responsibility, *assuming only* (A_1) (or (A'_1) whenever relevant).

Notwithstanding, both shared responsibility and rule consequentialism yield a strictly positive perceived induced externality for any prices p_b and p_g, and thus in particular even for $p_b < c_b$: $e^\dagger_{rc} > 0$, and $e^\dagger_{sr} > 0$. Hence, neutrality results stemming from $e^\dagger$ being nil whenever $[k_b + k_g + h_g(c_b)]c_b > 1$, fail. Moreover, as the perceived induced externality of investing in a polluting technology (i.e. any technology other than the green one) remains strictly positive, the moral pecking order over technologies always puts the green technology first.

[66]Namely, letting p_b^* be such that $p_b^*(k_b + i_b(p_b^*)) = 1$,

$$\frac{d(p_b e^\dagger_{rc})}{dp_b} = e^\dagger_{rc} = e[k_b + i_b(p_b^*)] > 0, \qquad \text{and} \qquad \frac{d(p_b e^\dagger_{sr})}{dp_b} = 0$$