Moral Preferences and the Invisible Hand

Rishabh Subramanian
University of Chicago

July 2026

Abstract

I study the efficiency of firms' production decisions when these decisions generate costs for consumers with moral concerns. Specifically, I study firms' choices between ethical and unethical production regimes when compared to an efficient benchmark. Firms only observe consumers' behavioral responses to moral concern, i.e., the change in their demand, while the efficient benchmark would value the level of moral harm itself. Two economies that differ in their level of moral harm faced by consumers may induce the same regime choice by firms even when the efficient benchmark would differ. I show that this leads to firms generically making inefficient regime choices when consumers display intermediate levels of moral concern, while they choose efficiently for high and low levels of moral concern. When moral concern rises, demand can become less elastic, letting firms generate greater markups in the unethical regime, which causes firms to persist with unethical production even when inefficient. In contrast, if the fall in demand due to increased moral concern is high enough, firms may choose to produce ethically even when the cost saving from unethical production makes switching to ethical production inefficient. A corrective Pigouvian tax or subsidy can restore efficiency in principle, but its rate cannot be identified from market data alone.

JEL: D11, D21, D61, L51
Keywords: moral consumption, production choice, inefficiency, market failure

1 Introduction

Consumers increasingly care not only about what they buy, but also about how it was produced. Was this shirt stitched by exploited workers? Was this chocolate harvested using child labor? Does buying from this firm support an oppressive regime? The effects of such concerns are not confined to private moral behavior; they increasingly shape consumers' purchasing decisions and, thereby, firms' production decisions. Consumers buy less from unethical firms, pay premiums for ethical certification, and reward brands that build their identity around cleaner production. For instance, after Russia's invasion of Ukraine, thousands of firms curtailed or exited Russian operations in response to public and consumer pressure (Sonnenfeld et al. (2022)). Brands such as Tony's Chocolonely and fair-trade coffee compete explicitly on the claim that their production regimes are morally superior.[1] The implicit promise of such moral consumption is powerful: if consumers care enough, decentralized markets will discipline firms into choosing better production.

At a first glance, the logic is straightforward. If consumers dislike unethical production, they reduce demand for goods produced unethically, or require a lower price to keep buying them. Firms then face lower demand, lower profits, and are incentivized to adopt ethical production. In this view, moral consumption allows decentralized markets to govern firms' production choices without direct regulation. However, this conclusion is not as clear-cut as it seems. Entire industries continue to use production regimes that consumers regard as unethical, even under consumer pressure. Moreover, ethical production is not always socially desirable. Although lower demand and moral harm to consumers make ethical production more socially appealing, if unethical production is substantially cheaper, switching to ethical production may destroy more surplus through higher costs than it saves in moral harm. This leads us to a fundamental question: can consumers' moral preferences, operating through decentralized markets, induce firms to choose the socially optimal production regime?

I show that the promise offered by moral consumption may not generally hold. Even if consumers are fully informed, perfectly rational, and their moral preferences are reflected in market demand, firms generically choose inefficient production regimes over a non-trivial range of moral concern. The mechanism is simple: firms respond to market signals while the efficient benchmark incorporates the level of moral harm as well. Firms observe the demand curve generated by moral concern and choose the regime that maximizes their profit but do not directly care about the level of moral harm faced by the consumers. In contrast, the efficient benchmark evaluates the total surplus generated by each regime, including the moral

[1] See https://www.forbes.com/sites/afdhelaziz/2026/07/07/how-tonys-chocolonely-is-reinventing-chocolate-as-a-force-for-good

cost consumers bear under unethical production. Moral consumption therefore disciplines firms, but it need not discipline them at the efficient point. As moral concern rises, the firm may switch to ethical production too late, continuing unethical production after ethical production has become socially efficient. It may also switch too early, abandoning unethical production even though its cost advantage still makes it socially efficient. The *central result* is therefore that moral consumption generically makes firms switch at the wrong level of moral concern.

I study this question in a model of production-regime choice. A firm chooses between an unethical and an ethical production regime. The unethical regime lowers marginal costs, while the ethical regime avoids the moral disutility consumers experience from consuming goods produced unethically. Consumers observe the production regime and choose how much to buy. As consumers' moral concern lowers their demand for the unethical good, the firm fully observes the market consequences of moral preferences through prices, quantities, and elasticities. The firm chooses the regime that yields higher profit. The efficient benchmark chooses the regime that yields higher total surplus. The model therefore gives moral consumption the strongest possible opportunity to discipline firms: consumers are fully informed, bear the moral cost directly, and express it through demand. The failure does not stem from ignorance, irrationality, or amoral individuals. Instead, it arises from the mechanism through which moral concern is translated into firm production incentives.

The result has direct implications for policy. Observing that firms respond to moral concerns is not sufficient to conclude that decentralized markets implement efficient production regimes. Although moral consumers can incentivize firms to switch production to an ethical regime, it also creates inefficiency that requires regulation. Moral demand and regulation are complements, not substitutes. However, this regulation is generally hard to determine. Market data reveal prices, quantities, demand responses, and profits. However, since moral harm itself is not observed, market data need not reveal the level of moral harm that determines the efficient switching threshold. A policymaker relying only on market behavior cannot generally determine how much moral harm is faced by consumers and, so, cannot determine whether firms are making efficient production choices or operating in the interval of threshold misalignment.

This paper contributes to a growing literature on how social and moral preferences enter consumer markets and influence firm behavior. Consumers may pay a premium for responsibly produced goods, avoid firms whose conduct they oppose, or participate in organized boycotts (Hart et al., 2024; Klein et al., 2004; Bartling et al., 2015; Kitzmueller and Shimshack, 2012). Firms may respond by supplying socially responsible attributes, bundling private goods with public-good provision, or changing their strategies in anticipation of con-

sumer and activist pressure (Baron, 2001; Bagnoli and Watts, 2003; Besley and Ghatak, 2007). Kaufmann et al. (2024) study how equilibrium feedback affects the choices of socially responsible consumers who seek to mitigate externalities. These papers generally show if and how social responsibility and morality affect consumption, and whether this can affect firm decisions. In contrast, this paper is concerned with the welfare consequences of moral behavior by consumers.

Broccardo et al. (2022) studies the relative effectiveness of exit versus voice strategies as mechanisms to induce socially optimal firm decisions when such decisions generate externalities. Socially responsible consumers are consequentialists: their incentive to boycott depends on the effect of their participation on the firm's behavior and, ultimately, on the external harm it produces. The potential inefficiency of exit consequently arises from a divergence between individuals' incentives to participate in a boycott and the social value of inducing the firm to adopt a clean technology. This paper instead holds the channel of consumer influence fixed and abstracts from externalities and strategic participation. Consumers behave as price takers and experience moral disutility from consuming under an unethical production regime, irrespective of whether their individual purchases are pivotal to the firm's choice. The relevant wedge is therefore not between the private and social value of joining a boycott, but between the moral preferences transmitted to the firm through demand and the full welfare consequences of the production regime.

Section 2 presents the model. Section 3 poses the central mechanism of inefficiency. Section 4 proves the existence of an inefficient interval. Section 5 discusses the policy implication. Section 6 extends the analysis from monopolies to Cournot competition, showing that the central result remains. Section 7 discusses empirical relevance. Section 8 concludes.

2 The Economy

Consider a market served by a representative firm with monopolistic pricing power.[2] There are two production regimes in which the firm can produce: one 'ethical' and one 'unethical' (as deemed by the consumers). In each production regime, the firm produces at a constant marginal cost and chooses the price that maximizes its profit, with unethical production being cheaper. A representative consumer observes the regime and the price and chooses how much to buy. Under unethical production, consumption of the good generates a moral disutility for the consumer. The firm is given a choice of which regime to produce in. Our primary question will be whether moral concern, expressed through demand, induces the firm to switch regimes, and whether the resulting regime choice is efficient.

[2]Section 6 extends the model to consider the firm being in Cournot competition

2.1 Preferences and Demand

The representative consumer receives quasilinear utility from consumption of the good, x, and a numéraire, y. Consumption of the good carries with it moral disutility, present when the firm produces the good unethically:

$$u(x, y; \delta) = U(x) - (1 - \delta)\gamma C(x) + y, \tag{1}$$

where $\gamma \geq 0$ represents the level of moral concern of the consumer and $\delta \in \{0, 1\}$ represents the regime in which the good is produced. Unethical production is indexed by $\delta = 0$ and ethical by $\delta = 1$. Moral concern being inherent to consumers' utility, i.e., deontological, and not dependent on the effects of the concern is based on the results of Hart et al. (2024). The consumer has a budget constraint, $px + y \leq m$, giving them bounded solutions, but the quasilinear utility makes all income effects zero. Let $x_0(p, \gamma)$ denote the demand function under unethical production and $x_1(p)$ the demand function under ethical production. When no confusion arises, equilibrium demands are denoted simply by x_δ.

Assumption 1 (Utility)**.** *$U(0) = 0$, $U' > 0$, $U'' < 0$, and U is thrice continuously differentiable. The ethical choke price is finite: $U'(0^+) = \bar{p} < \infty$.*

Assumption 2 (Moral Disutility)**.** *$C(0) > 0$, $C'' > 0$, and C is thrice continuously differentiable. The two regimes share the choke price: $C'(0^+) = 0$; together with $C'' > 0$ this gives $C' > 0$ for $x > 0$.*

We set $U(0) = 0$ for simplicity. $C(0) > 0$ represents moral concern for unethical production itself, independent of the amount of the good consumed. The consumer's first-order conditions for the demand functions are

$$U'(x_1(p)) = p, \qquad U'(x_0(p, \gamma)) - \gamma C'(x_0(p, \gamma)) = p. \tag{2}$$

Moral concern only enters demand through the derivatives of moral cost: $C'(x)$ lowers willingness to pay and $C''(x)$ changes its slope. The level of the moral cost, $C(x)$, does not enter demand by itself at all. Therefore, two moral-cost functions differing by a constant create the same demand; a consumer facing $C(x)$ and one facing $C(x) + 100$ demand the same amount at every price. This is the first appearance of an asymmetry we will return to: the firm, which only observes demand, cares about $C'(x)$ and higher derivatives, while welfare will also incorporate the level $C(x)$.

The first-order conditions show us that a consumer who cares about morality treats each unit bought under unethical production as carrying a hidden surcharge equal to its marginal

moral cost, $\gamma C'(x)$. This surcharge enters the consumer's first-order condition exactly as a higher price would, so at any given posted price the morally concerned consumer buys weakly less under unethical production, and strictly less as soon as $\gamma > 0$. An 'amoral' consumer, with $\gamma = 0$, demands the same in both states. We can formalize this as follows.

Lemma 1 (Abstention Effect)**.** *For any common interior price $p \in (0, \bar{p})$, we have: (1) $x_0(p, \gamma) \leq x_1(p)$ with equality only if $\gamma = 0$, and (2) $\dfrac{\partial x_0}{\partial \gamma} < 0$ and $\dfrac{\partial x_1}{\partial \gamma} = 0$.*

Everything the firm learns about consumers' moral concern, it learns from this Abstention Effect, or the change in the demand. Two features of that movement will matter later. The first is the downward shift in quantity, which is the consumers' intended punishment of unethical production. The second is a change in the slope of demand, visible through the price derivatives of demand:

$$x_0'(p, \gamma) = \frac{1}{U''(x_0) - \gamma C''(x_0)}, \quad x_1'(p) = \frac{1}{U''(x_1)}. \tag{3}$$

This is where C'' first enters the model: it does not change the level of moral harm or demand itself, but it changes the shape of the demand curve the firm prices against, and so it matters when the firm has pricing power.

Note that the comparison in the lemma holds price fixed; equilibrium demands need not be ordered this way, as the firm generally sets a different price in each regime. Because of this ambiguity caused by the difference in price between the regimes, it is generally unclear whether the consumers are better off in the ethical regime. So, in order to analyze this, we consider the equilibrium consumer surplus, given by

$$CS_\delta := U(x_\delta) - (1 - \delta)\gamma C(x_\delta) - p_\delta x_\delta. \tag{4}$$

This will become important later in our welfare analysis of the firm's production decision.

2.2 Production, Pricing, and Private Choice

The firm chooses its price in each regime. Its profit objective is $\Pi_\delta(p) := (p - c_\delta)x_\delta(p)$, where c_δ denotes the marginal cost in regime δ. We have $c_0 < c_1 < \bar{p}$, which ensures that both regimes are profitable. The maximized profit in each regime is given by

$$\pi_\delta := \max_{p \in [c_\delta, \bar{p}]} \Pi_\delta(p). \tag{5}$$

Prices below marginal cost yield negative profit, and prices above the choke price generate zero demand, so the effective price choice set is just $[c_\delta, \bar{p}]$. Let $\Delta\pi$ denote the difference in the maximized profits between the ethical and unethical regime:

$$\Delta\pi := \pi_1 - \pi_0. \tag{6}$$

The firm chooses to produce in the ethical regime if and only if it is able to produce greater profits by switching; that is, if and only if $\Delta\pi \geq 0$.

Assumption 3 (Profit Regularity)**.** *In both regimes, $\Pi_\delta(p)$ is strictly quasi-concave in p on $(c_\delta, \bar{p})$, guaranteeing a unique optimal price, which is continuous in all primitives.*

Let p_δ denote the firm's optimal price in regime δ. The firm's markup is given by

$$\mu_\delta := p_\delta - c_\delta = \frac{x_\delta(p_\delta)}{|x'_\delta(p_\delta)|}. \tag{7}$$

Lemma 2 (Profit Monotonicity)**.** *Unethical profit is strictly decreasing in moral concern, while ethical profit is independent of moral concern. That is $\pi'_0(\gamma) < 0$ and $\pi'_1(\gamma) = 0$; consequently, $\Delta\pi'(\gamma) > 0$.*

This lemma is a consequence of Lemma 1. An increase in the level of moral concern decreases demand at all prices in the unethical regime. Therefore, the unethical profit objective, $\Pi_0(p)$, falls everywhere, making the maximized unethical profit fall as well. This monotonicity is central to the efficacy of moral concern in creating change in the market. Unethical profit *always* decreases with increases in moral concern in the economy, therefore making switching to the ethical regime more favorable.

2.3 Surplus and the Efficient Benchmark

Adding consumer surplus and profit yields the total surplus in the economy:

$$W_\delta := CS_\delta + \pi_\delta = U(x_\delta) - (1-\delta)\gamma C(x_\delta) - c_\delta x_\delta. \tag{8}$$

The payment $p_\delta x_\delta$ cancels because consumer expenditure is firm revenue. Let $\Delta W := W_1 - W_0$ represent the difference in welfare between the regimes. Choosing to produce in the ethical regime is efficient if and only if the total surplus generated in the ethical regime is higher than that in the unethical regime; that is, if and only if $\Delta W \geq 0$. Taking differences

of our welfare expression across regimes, we get

$$\Delta W = \underbrace{[U(x_1) - U(x_0)]}_{\text{consumption distortion}} + \underbrace{\gamma C(x_0)}_{\text{moral gain}} - \underbrace{[c_1 x_1 - c_0 x_0]}_{\text{cost increase}}. \tag{9}$$

Lemma 3 (Welfare Monotonicity)**.** *Unethical welfare is strictly decreasing in moral concern, while ethical welfare is independent of moral concern. That is,* $W_0'(\gamma) < 0$ *and* $W_1'(\gamma) = 0$*; consequently,* $\Delta W'(\gamma) > 0$*.*

Welfare, unlike profit, counts the moral harm consumers bear directly. Differentiating unethical welfare along the equilibrium path, and using the consumer's first-order condition (2) to collapse terms, we get

$$\frac{dW_0}{d\gamma} = (p_0 - c_0)\frac{dx_0}{d\gamma} - C(x_0) \tag{10}$$

Both terms are negative. Higher moral concern shrinks the equilibrium quantity, $dx_0/d\gamma < 0$, and since the firm prices above marginal cost, each unit lost destroys surplus equal to the markup. Separately, the consumers who continue to buy bear more harm. One channel cuts the units and the other raises the harm per unit, so neither can offset the other.

Both profit and welfare therefore increase with increases in moral concern. So, in order to understand the behavior of their curves, we need to pin down their signs at the edges of moral concern. At $\gamma = 0$, production in the ethical regime faces a higher cost but faces the same demand curve. Therefore, profit in the ethical regime is strictly less than that in the unethical regime. Similarly, the higher cost in the ethical regime causes its price to be strictly higher, making the consumers worse off. So, consumer surplus in the ethical regime is strictly less than that in the unethical regime. With $\Delta W = \Delta\pi + \Delta CS$, this gives us $\Delta W(0) < \Delta\pi(0) < 0$. In the limit as $\gamma \to \infty$, by contrast, consumption in the unethical regime goes to zero, making unethical profit go to zero as well. So, both the profit and welfare comparisons become positive. In particular, because of the inherent moral harm caused by production in the unethical regime, $C(0) > 0$, ΔCS also becomes positive. Therefore, as $\gamma \to \infty$, $0 < \Delta\pi < \Delta W$. Both the comparisons therefore cross zero at some point, and the points at which they do determine whether the firm's regime choice is efficient, which is investigated in Section 4.

3 The Profit–Welfare Mismatch

The firm and the efficient benchmark rank the same two regimes against different objectives. The firm's payoff is its profit, in which the moral cost C does not appear. In contrast, the efficient benchmark's objective is total surplus, in which C appears directly. This is the

central mechanism that can cause the firm's decisions to differ from those of the efficient benchmark.

Moral concern reaches the firm only through demand. It lowers willingness to pay through the marginal moral cost $\gamma C'(x)$ and steepens the demand curve through $\gamma C''(x)$. The markup carries both of these effects:

$$\mu_0 = \Psi(x_0) + \gamma x_0 C''(x_0), \tag{11}$$

where $\Psi(x) := -xU''(x)$ is the baseline monopoly markup function. Consumers punish unethical production by withdrawing their purchasing; the firm only feels that withdrawal, and acts in light of it.

Proposition 1 (Profit–Welfare Mismatch)**.** *Under Assumptions 1–3, moral concern enters the firm's private regime comparison only through the demand curve it induces and the profit generated by that demand curve. The direct moral harm $\gamma C(x_0)$ enters the welfare comparison but does not enter the firm's profit objective. Hence the private comparison $\Delta\pi$ and the welfare comparison ΔW generally respond to different features of the same moral preference.*

Consider adding a constant to the moral cost function. Every derivative is unchanged, so demand, price, markup, and profit are unchanged, making the firm face an identical problem and make identical choices. But the harm consumers bear has shifted, and so the social value of unethical production has shifted as well. Two economies can therefore agree in every market statistic—same demand curve, same prices, same quantities, same profits—but disagree about whether unethical production is efficient. I call this invariance the *Missing Levels Channel*: the market transmits the derivatives of moral cost to the firm but discards the level, while welfare includes the level as well.

This channel is the central mechanism for inefficiency in moral economies. Profit maximizing firms only care about the demand changes that occur due to moral concerns; they do not truly care about the moral harm caused by their production. So, even though moral concern by consumers decreases firms' profits, incentivizing them to switch regimes, they might not do so correctly because they do not internalize the moral harm itself.

4 The Inefficiency of Firm Choices

The mismatch between profit and welfare identified in the previous section does not mean that the firm always makes the inefficient production choice, and, as we will see, it indeed does not. This section therefore analyzes the consequences of the Missing Levels channel, studying if and how firms can make the wrong production regime choice.

Let γ_π solve $\Delta\pi(\gamma_\pi) = 0$ and γ_W solve $\Delta W(\gamma_W) = 0$. These are the thresholds for regime choice change. The firm produces ethically for all levels of moral concern above γ_π and ethical production is efficient for all levels of moral concern above γ_W.

Lemma 4 (Existence of Thresholds)**.** *Under Assumptions 1–3, the thresholds γ_π and γ_W are both positive, finite, and unique.*

Because $\Delta\pi$ and ΔW are strictly increasing, neither can return to zero once it has left it, so each threshold defines a single switching point rather than a set of them. Therefore, the differences between firm and efficient decisions can be understood simply by looking at their thresholds. Below both thresholds, profit and welfare favor unethical production. Above both, profit and welfare favor ethical production. Therefore, any conflict between firm and efficient choice happens between the thresholds. That is, whenever the private and efficient thresholds differ, their decisions conflict for all levels of moral concern between them.

Proposition 2 (Threshold Errors)**.** *If $\gamma_\pi \neq \gamma_W$, the firm's regime choice is inefficient precisely on the open interval between the thresholds, and efficient everywhere else. Two cases arise:*

1. *If $\gamma_W < \gamma_\pi$, the firm* switches too late *on (γ_W, γ_π): ethical production is more efficient, but the firm still produces unethically.*

2. *If $\gamma_\pi < \gamma_W$, the firm* switches too early *on (γ_π, γ_W): the firm produces ethically, but unethical production is more efficient.*

The right and left tails of γ induce efficient regime choices.

So, firm regime choices can be inefficient if the switching thresholds in fact are not equal. However, it could be that the thresholds are often, or even always, equal. This is not the case. Almost every economy has differing switching thresholds and therefore an interval of inefficient choice.

Theorem 1 (Generic Threshold Separation)**.** *Under any smooth distribution over the primitives that define the economy, the probability that a randomly drawn economy has equal private and efficient thresholds is 0:*

$$\mathbb{P}(\gamma_\pi = \gamma_W) = 0.$$

Consequently, with probability one there is a non-empty open interval of moral concern on which the firm's regime choice is inefficient.

This theorem is a formal consequence of the Missing Levels Channel. Because the firm and efficient objectives differ on the incorporation of moral harm, their choices are likely to differ. Specifically, if we were to randomly draw economies from a distribution that represents the plausibility of economies, a condition that would make firm and efficient choices align would be too restrictive to be plausible. We assume that the distribution over economies must be smooth as to not eliminate large sections of the economy. If an economy where unethical production has a marginal cost of \$5 and ethical production has one of \$10 is plausible, one with marginal costs \$5.01 and \$10.01 must be similarly plausible. So, small perturbations shouldn't cause massive changes in the plausibility of an economy, and so the distribution should be smooth.

Therefore, the inefficiency of firm production decisions is not a coincidental oddity but is generally a full interval of moral concern across which the firm produces in the less efficient regime. The direction of this inefficiency—whether the firm switches too early or too late, as shown in Proposition 2—is determined by consumer surplus. Since $\Delta W = \Delta\pi + \Delta CS$, evaluating this at the profit threshold gives us:

$$\Delta CS(\gamma_\pi) < 0 \quad \iff \quad \gamma_\pi < \gamma_W. \tag{12}$$

So, the direction of inefficiency is determined by whether consumers are better off in the ethical regime when the firm switches. However, there is no general force in the model that signs this quantity. If the firm wishes to switch to the ethical regime because they can sell at a much higher price there, consumers might be better off in the unethical regime. In contrast, if the firm wishes to stay in the unethical regime due to the lower costs, consumers' moral concerns might make them better off in the ethical regime. Therefore both cases of inefficiency are plausible. Figure 1 shows these two cases.

The too-late case has an interesting mechanism that causes it. When consumers' moral concern rises, the consumers who exit can be the more elastic ones, making the ones who stay less elastic. Consumer exit can therefore leave a less elastic residual demand from our representative consumer. So, a firm with market power can respond to heightened moral concern by raising its markup. The higher markup cushions unethical profit against the very concern that was meant to erode it, so unethical profit reduces slower than intended and the firm continues producing unethically well past the point society would have it stop. This is exactly the too-late case. Moral concern, by pruning away the price-sensitive consumers, hands the firm the market power to withstand it. I call this the *Elasticity Trap*: rather than accelerating the transition to ethical production, heightened moral concern from consumers can delay it.

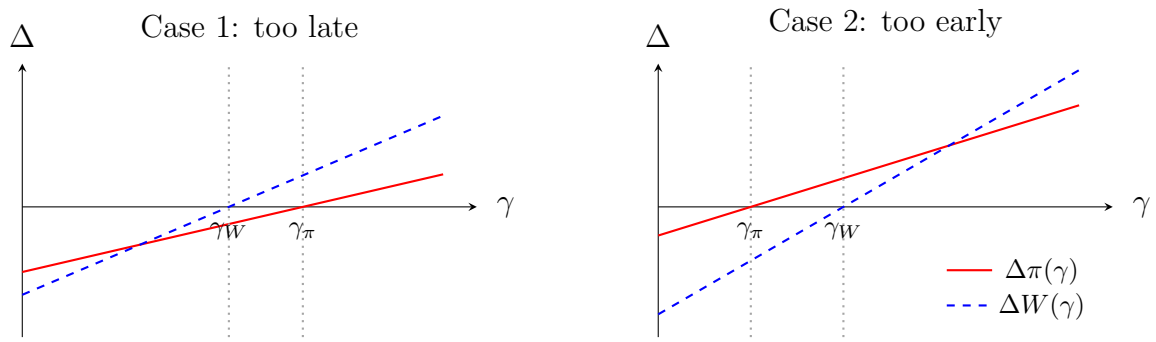

Figure 1: Private and efficient switching as moral concern varies. The solid red line is the profit difference $\Delta\pi(\gamma)$, which crosses zero once at γ_π; the dashed blue line is the welfare difference $\Delta W(\gamma)$, which crosses zero once at γ_W. Both are strictly increasing. The two cases are distinguished by the order of the thresholds. In Case 1 the welfare threshold comes first ($\gamma_W < \gamma_\pi$), so on (γ_W, γ_π) ethical production is efficient while the firm still produces unethically—it switches too late. In Case 2 the profit threshold comes first ($\gamma_\pi < \gamma_W$), so on (γ_π, γ_W) the firm produces ethically while unethical production is still efficient—it switches too early.

5 Can Firm Choices be Efficient?

The previous section and Theorem 1 show that the moral market is almost always inefficient over some interval of moral concern. They do not say where that interval lies. So, a natural hope for a policymaker would be that they live in the efficient part of the economy and not in the inefficient interval. Another hope would be that policy could repair the damage if the economy is in the inefficient interval by detecting it and correcting it where it occurs. This section shows that the hope runs into an obstacle built into the mechanism itself.

5.1 Unobservability of the Inefficient Region

Generic inefficiency would not be a problem for policy if a policymaker could observe which markets lie in the inefficient region and intervene with perfectly tuned policy there. However, they cannot. The Missing Levels channel, which separates the firm's threshold from the planner's, also masks the level of moral harm from market data. The inefficiency is not only generic but, in a sense, invisible, and so it is hard to form policy to mitigate the inefficiency.

The firm's threshold depends on demand and cost. Both are visible: they are recovered from prices, quantities, and the response of demand to price. The planner's threshold depends in addition on the level of moral harm, the total moral cost consumers bear under unethical production. Moral concern reaches the firm only through willingness to pay, so the market records how moral concern reshapes demand but not the harm behind it. Markets

with identical prices, quantities, and demand responses can carry different levels of moral harm, and therefore different efficient thresholds.

This is the wedge of the original mechanism, restated as an identification problem. The firm's threshold is built from observables; the planner's is built partly from an object the market conceals. A policy premised on lying in an aligned tail cannot be verified, because the data consistent with that premise are equally consistent with its negation.

5.2 Pigouvian Policy

The robust objective is therefore not to confirm efficiency but to eliminate the inefficient region: to align the firm's threshold with the planner's so that the firm chooses efficiently at every level of moral concern. The natural instrument is a corrective tax or subsidy on unethical production, which shifts the firm's regime payoff and moves its switching point.

Such an instrument always exists. The firm's regime choice turns on the sign of $\Delta\pi$; a transfer t levied on unethical production replaces this comparison with $\Delta\pi + t$ and moves the firm's threshold to the γ at which $\Delta\pi(\gamma) = -t$. Because $\Delta\pi$ is continuous and strictly increasing from $\Delta\pi(0) < 0$ to $\pi_1 > 0$, some t places that threshold exactly at γ_W; the aligning rate is

$$t^* = -\Delta\pi(\gamma_W) = \Delta CS(\gamma_W), \tag{13}$$

the consumer-surplus wedge at the welfare threshold. Because the transfer conditions on the regime rather than on output, it leaves within-regime prices, quantities, and welfare untouched and moves only the switch, so γ_W is unchanged and the firm switches exactly where switching becomes efficient. The sign of t^* is the direction of the error: by (12), $t^* > 0$ is a tax when the firm switches too late ($\gamma_W < \gamma_\pi$) and $t^* < 0$ is a subsidy when it switches too early. First-best regime choice is therefore always implementable in principle.

Proposition 3 (Pigouvian Non-Identification)**.** *For every economy there is a corrective tax or subsidy on unethical production that aligns the firm's switch with the efficient threshold—a tax when the firm switches too late, a subsidy when it switches too early—with aligning rate* $t^* = \Delta CS(\gamma_W)$*. This rate is not identified by market data: prices, quantities, and demand responses fix the firm's threshold* γ_π *but not* γ_W*, which depends on the moral-harm level* $C(x_0)$*. Two economies with identical market data can require aligning rates of different sign and size.*

The failure is therefore not that no corrective instrument exists, nor merely that markets and regulation act on different margins. It is that the correct instrument cannot be recovered from what the market reveals. The aligning rate $t^* = \Delta CS(\gamma_W)$ is built from the welfare

threshold, and γ_W turns on the level of moral harm—the one object the market discards. A policymaker who sees prices, quantities, demand responses, and profits sees everything that fixes γ_π and nothing that fixes γ_W, so the sign and size of the needed correction are invisible even though the correction itself is available. Moral demand and regulation are complements, not substitutes: were the level of moral harm known from outside the market, the gap could be closed exactly, but market data never deliver it.

6 Extension to Cournot Competition

This section extends the model to consider the firm being in Cournot competition to show that the identified inefficiency does not rely on the firm being a monopoly. Consider N identical firms producing a homogeneous good and competing in quantities. The conclusion survives. Under oligopoly the planner's side of the model retains the monotone structure it had under monopoly, while the industry's side loses it. The efficient benchmark is therefore almost unchanged, the industry's behavior becomes considerably more erratic, and the inefficiency remains generic. Because the firms are identical, the production regime is an industry-wide choice: we compare the symmetric equilibrium of the all-ethical industry with the symmetric equilibrium of the all-unethical industry.

Aggregate output is $X = \sum_{i=1}^{N} q_i$. From the consumer's first-order condition, inverse demand in regime δ is $P_\delta(X, \gamma) = U'(X) - (1 - \delta)\gamma C'(X)$; we write $P(X, \gamma) := U'(X) - \gamma C'(X)$ for the unethical regime. Each firm chooses q_i to maximize $(P_\delta - c_\delta)q_i$ taking rivals' quantities as given.

Assumption 4 (Cournot Regularity)**.** *In both regimes and at all relevant (X, γ): (i) a symmetric Cournot equilibrium exists, is unique, and is interior, with aggregate output continuously differentiable in all primitives; (ii) industry marginal revenue is strictly decreasing,* $2P_X + XP_{XX} < 0$.

Assumption 4 is the Cournot counterpart of Assumption 3 and asks no more of the model than that one does: it delivers a well-defined equilibrium that moves smoothly with the primitives, and nothing else. It is not an additional restriction on preferences. Since $P_X = U'' - \gamma C'' < 0$, condition (ii) already implies the stability condition

$$(N+1)P_X + XP_{XX} \;=\; \big(2P_X + XP_{XX}\big) + (N-1)P_X \;<\; 0$$

for every $N \geq 1$, so equilibrium comparative statics are well behaved without a separate assumption. In particular, nothing is imposed on the curvature of the moral cost. The consequences of leaving that curvature free are the substance of this section.

The symmetric equilibrium in regime δ satisfies

$$P_\delta(X_\delta, \gamma) + \frac{X_\delta}{N} \frac{\partial P_\delta}{\partial X}(X_\delta, \gamma) \;=\; c_\delta, \tag{14}$$

with per-firm markup $(X_\delta/N)\,|\partial P_\delta/\partial X| > 0$ for every finite N. Per-firm maximized profit is $\pi_\delta(\gamma)$, total surplus is $W_\delta(\gamma) = U(X_\delta) - (1-\delta)\gamma C(X_\delta) - c_\delta X_\delta$, and $\Delta\pi := \pi_1 - \pi_0$, $\Delta W := W_1 - W_0$ as before.

6.1 Monotone Welfare, Non-Monotone Profit

Proposition 4 (Asymmetric Robustness of Monotonicity)**.** *Under Assumptions 1–2 and 4, for every finite N: equilibrium unethical output is strictly decreasing in moral concern, $dX_0/d\gamma < 0$; unethical welfare is strictly decreasing in moral concern and ethical welfare is independent of it, $W_0'(\gamma) < 0$ and $W_1'(\gamma) = 0$, so ΔW is strictly increasing exactly as in Lemma 3. Per-firm unethical profit satisfies*

$$\pi_0'(\gamma) \;=\; \frac{X_0}{N}\Big[\; \underbrace{-\,C'(X_0)}_{\text{direct erosion}} \;+\; \underbrace{\tfrac{N-1}{N}\,P_X\,\tfrac{dX_0}{d\gamma}}_{\text{strategic cushioning}} \;\Big],$$

whose sign is ambiguous for every $N \geq 2$. Hence $\Delta\pi$ need not be monotone in γ.

Welfare survives intact, and it does so without any restriction beyond regularity. The two channels of Lemma 3 are unchanged: output is lost at a positive markup, and the consumers who continue to buy bear greater harm. Both are negative at every market structure and every finite N, so the planner's problem under oligopoly is the problem it was under monopoly. The efficient benchmark still flips once, at a single threshold, and everything the monopoly analysis said about it carries over.

Profit does not survive. The strategic term is new and it is positive: rising moral concern contracts each rival's output, and since $P_X < 0$, that contraction raises the residual price a firm faces. When the moral cost is sharply convex, the contraction is large and this cushioning can outweigh the direct erosion of the firm's own demand, so unethical profit can *rise* with moral concern. The channel is absent at $N = 1$, where the bracket collapses to $-C'(X_0) < 0$; this is why the monopoly model delivered monotone profit with no restriction on C''. Non-monotonicity is strictly a phenomenon of oligopoly, and it is what a curvature condition on the moral cost would have assumed away.

6.2 First and Last Crossings

The loss of monotonicity costs the industry its single switching point. $\Delta\pi$ is still continuous with $\Delta\pi(0) < 0$ and $\Delta\pi(\gamma) \to \pi_1 > 0$, so its zero set is non-empty and compact, but it need not be a singleton. Define the first and last crossings

$$\underline{\gamma}_\pi := \min\{\gamma \geq 0 : \Delta\pi(\gamma) = 0\}, \qquad \overline{\gamma}_\pi := \max\{\gamma \geq 0 : \Delta\pi(\gamma) = 0\}, \tag{15}$$

with $\underline{\gamma}_\pi \leq \overline{\gamma}_\pi$, and equality exactly when the industry switches once and for all. Below $\underline{\gamma}_\pi$ the industry produces unethically and above $\overline{\gamma}_\pi$ it produces ethically; on $(\underline{\gamma}_\pi, \overline{\gamma}_\pi)$ it may switch, revert, and switch again. Welfare, still monotone, continues to deliver the single threshold γ_W.

Proposition 5 (Thresholds and Inefficiency under Cournot)**.** *Under Assumptions 1–2 and 4, the thresholds in (15) and γ_W are well defined and finite, and:*

1. The tails remain aligned. *For $\gamma < \min\{\gamma_W, \underline{\gamma}_\pi\}$ and for $\gamma > \max\{\gamma_W, \overline{\gamma}_\pi\}$, the industry's regime choice is efficient.*
2. *If $\gamma_W < \underline{\gamma}_\pi$, the industry switches too late on the non-empty interval $(\gamma_W, \underline{\gamma}_\pi)$.*
3. *If $\gamma_W > \overline{\gamma}_\pi$, the industry switches too early on the non-empty interval $(\overline{\gamma}_\pi, \gamma_W)$.*
4. *If $\underline{\gamma}_\pi \leq \gamma_W \leq \overline{\gamma}_\pi$, the regime choice is inefficient on $\{\gamma : \Delta\pi(\gamma)$ and $\Delta W(\gamma)$ differ in sign$\}$, a subset of $(\underline{\gamma}_\pi, \overline{\gamma}_\pi)$ that alternates between episodes of premature and belated switching.*

The industry's regime choice is efficient at every level of moral concern if and only if $\underline{\gamma}_\pi = \overline{\gamma}_\pi = \gamma_W$.

The last statement is the one that matters, and it is more demanding than its monopoly counterpart. Under monopoly, efficiency everywhere required two numbers to agree. Here it requires the industry to cross exactly once *and* to cross at γ_W: the profit comparison must be single-crossing, and its unique crossing must land on the welfare threshold.

Two features of the oligopoly economy deserve emphasis. The first is that the planner and the industry come apart in kind, not only in location. The efficient benchmark still behaves simply: ethical production becomes efficient once and stays efficient, and the comparative static "more moral concern makes ethical production more attractive" holds without qualification. The industry does not. Because unethical profit can rise with γ, an industry that has switched to ethical production may revert to unethical production as consumers come to care *more*. Increasing moral concern therefore need not make firms more prone to switch,

and observing an industry switch at one level of moral concern gives no assurance that it will remain switched at a higher one. The monotone discipline that moral demand exerted on a monopolist is not a general property of moral markets; it is a property of the single-firm case.

The second is that the tails still align. Very low and very high moral concern still produce efficient choices, for the same reason as under monopoly: at the extremes the profit and welfare comparisons agree in sign regardless of what happens in between. The inefficiency remains an interior phenomenon. What oligopoly adds is that the interior is no longer a single interval of error but can be a union of alternating too-early and too-late episodes.

6.3 Genericity

Non-monotone profit does not rescue the industry from inefficiency. It merely changes the shape of the inefficient region.

Proposition 6 (Genericity of Inefficiency under Cournot)**.** *Let ϑ denote a finite-dimensional parameterization of the economy's admissible primitives (U, C, c_0, c_1), with the additive level of C included as a free coordinate. If ϑ is drawn from any probability distribution absolutely continuous with respect to Lebesgue measure, then $\mathbb{P}\big(\underline{\gamma}_\pi = \gamma_W\big) = 0$. Consequently, with probability one there is a non-empty open set of moral concern on which the industry's regime choice is inefficient.*

The argument is the one behind Theorem 1, applied to the first crossing rather than to the unique one. The additive level of moral harm leaves the industry's profit comparison and all of its crossings unchanged, but shifts the welfare comparison. Conditional on all other primitives, at most one level can make the first private crossing coincide with the welfare threshold. Non-monotone profit therefore changes the shape of the inefficient region, not the genericity of its existence.

The implementation failure is therefore not driven by monopoly. It survives symmetric Cournot competition at any finite N, because its source, the omission of the moral-harm level from the private objective, is untouched by the number of firms. Monopoly is not essential for the central mechanism; market power is. What oligopoly changes is the industry's response to moral concern, which becomes non-monotone and can reverse, while the efficient benchmark it is failing to track stays exactly as simple as before.

7 Empirical Relevance

The model rests on three empirical premises: consumers may attach moral value directly to firms' production practices; these preferences affect purchasing behavior; and moral attributes may alter not only the level of demand but also the way consumers respond to prices. This section reviews evidence supporting those premises, relates several existing empirical findings to the model's mechanisms, and then states predictions that are distinctive to the framework.

7.1 Empirical Support for the Model's Primitives

The assumption that moral concern enters utility directly is motivated most closely by Hart et al. (2024). They find that consumers remain willing to sanction firms associated with objectionable conduct even when experimentally varied beliefs suggest that the sanction is unlikely to affect the firm. This relative insensitivity to perceived impact is consistent with a deontological motive: consumers may avoid a firm because purchasing from it is itself regarded as morally objectionable, rather than only because abstention is expected to change the firm's conduct. The evidence is also consistent with moral disutility attaching partly to the firm's production regime itself, rather than solely to the consequences of an individual's marginal purchase. Related evidence in Kaufmann et al. (2024) identifies a non-trivial group of consumers who value socially responsible actions even when those actions have no aggregate effect, although consequentialist motivations remain more common in their survey.

A separate body of evidence shows that moral and political concerns generate meaningful purchasing responses. Chavis and Leslie (2009) document a substantial decline in French wine sales in the United States during the consumer boycott associated with opposition to France's position on the Iraq War. Pandya and Venkatesan (2014) find that the same conflict reduced the market shares of French-sounding American brands, suggesting that consumers responded to the perceived political association of the products rather than only to changes in their physical characteristics. In experimental product markets, Bartling et al. (2015) find that consumers are willing to pay more for products produced under socially responsible conditions. Together, these findings support the model's basic premise that information about firms' conduct can shift demand even when the product itself is otherwise unchanged.

The curvature assumption requires more than a demand-level response: moral attributes must also be capable of changing consumers' price sensitivity. Evidence from field experiments suggests that this is empirically plausible. Hainmueller et al. (2015) find that Fair Trade labels increase coffee sales and that the response to accompanying price increases dif-

fers sharply across products. Related evidence indicates that fair-labor labels can sustain increased demand even when accompanied by price premia (Hiscox and Smyth, 2011). These studies do not directly identify the model's curvature condition, since ethical labels may also affect perceived quality, product differentiation, or consumer composition. They nevertheless show that moral attributes can interact with price responses rather than operating as uniform parallel shifts in demand.

7.2 Existing Findings Consistent with the Mechanism

Several empirical findings are consistent with the model's distinction between consumer response and firm response. Consumer pressure can materially affect demand without translating proportionally into changes in production practices. In experimental markets, Pigors and Rockenbach (2016) find that socially responsible production affects consumers' purchasing decisions and firms' profits, but that its emergence depends strongly on the surrounding market structure. Responsible production is difficult to sustain under monopoly but can become profitable when suppliers compete for socially responsible consumers. Although their environment differs from the present model, the results support the broader mechanism that the same consumer preferences can create different production incentives depending on how firms price and compete.

Evidence from actual boycotts similarly indicates that consumer opposition does not translate mechanically into firm reform. King (2008) finds that firms' concessions to boycott campaigns depend substantially on media amplification and the target's existing reputational position. This evidence is often interpreted through participation, publicity, and reputational channels. The present model adds a complementary pricing channel. Even after moral concern has entered demand, its effect on profit and production choice depends on the demand curve against which the firm optimizes. A visible consumer response may therefore produce a muted, delayed, or non-monotone production response.

The framework also changes how empirical success is interpreted. Studies of moral consumption and boycotts frequently treat a change in firm conduct as the relevant measure of effectiveness. In the model, however, behavioral influence and welfare improvement are distinct. A firm that retains an unethical regime need not be making an inefficient choice when the cost advantage of that regime remains sufficiently large. Conversely, inducing a switch to ethical production need not improve welfare if the additional production cost exceeds the moral benefit. Observed persistence is therefore not sufficient evidence of market failure, and observed reform is not sufficient evidence of market success.

7.3 Novel Predictions

The first prediction is that stronger moral concern may produce falling quantities together with rising prices while the firm remains in the unethical regime. Moral concern lowers demand directly, but it can also change the elasticity of the residual demand faced by a firm with market power. If the latter effect is sufficiently strong, the firm may raise its price even as sales decline. This joint pattern is not unique to the mechanism, since cost or supply shocks could generate a similar response. Conditional on unchanged production costs, product quality, and production regime, however, it is consistent with moral concern being translated partly into a higher markup rather than solely into lower profit.

Second, moral concern can reduce consumer welfare even before it induces any observable production reform. Consumers who continue purchasing under the unethical regime bear greater moral disutility, and they may also face a higher price. A boycott or moral-consumption campaign can therefore appear ineffective when evaluated only by whether the firm changes its production regime, while still having substantial effects on the welfare of consumers who remain in the market.

Third, the relationship between moral concern and firm behavior need not be monotone under competition. In the Cournot extension, contraction by rival firms can cushion each firm's residual demand. An industry may therefore switch to ethical production, revert to unethical production as moral concern increases further, and switch again at still higher levels of concern. Stronger consumer opposition need not continuously strengthen firms' incentives to reform.

Finally, the magnitude of the observed demand response is not sufficient to rank the welfare performance of moral markets. A larger reduction in demand may accelerate an efficient switch, induce a premature switch, or be partly absorbed through the firm's pricing response. Empirical measures of boycott participation, sales losses, or willingness to pay therefore capture the strength of consumer influence but not, by themselves, whether that influence moves production choices closer to the efficient benchmark.

8 Conclusion

Moral preferences can discipline firms, but they need not do so efficiently. The reason is not that consumers fail to care. Even when all consumers are moral, their concern reaches the firm through demand. A price-setting firm translates that demand into prices and markups before it translates it into a production-regime switch.

The central result is a threshold error. Moral concern varies continuously, but the firm

changes regimes discretely. The private decision threshold is governed by profit, while the efficient benchmark is governed by total surplus. Because profit depends on moral curvature and welfare depends on moral levels, these thresholds need not coincide. The firm may switch too early or too late.

The invisible hand is therefore incomplete in moral markets. It moves firms away from unethical production if moral concern is sufficiently high. However, firm choices induced by moral concern need not be efficient. This conclusion should not be interpreted as an argument against moral consumption or against decentralized markets. Moral preferences remain an important force for disciplining firms and shifting production toward more ethical practices. Rather, the analysis identifies a limit to what moral consumption can achieve on its own. When firms observe only the market response generated by consumers' moral concerns, decentralized markets cannot generally be expected to implement the efficient production regime. Markets disciplining firms through consumer moral concern and socially efficient production choice generally differ.

References

Bagnoli, M. and Watts, S. G. (2003). Selling to socially responsible consumers: Competition and the private provision of public goods. *Journal of Economics & Management Strategy*, 12(3):419–445.

Baron, D. P. (2001). Private politics, corporate social responsibility, and integrated strategy. *Journal of economics & management strategy*, 10(1):7–45.

Bartling, B., Weber, R. A., and Yao, L. (2015). Do markets erode social responsibility? *The Quarterly Journal of Economics*, 130(1):219–266.

Besley, T. and Ghatak, M. (2007). Retailing public goods: The economics of corporate social responsibility. *Journal of public Economics*, 91(9):1645–1663.

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

Chavis, L. and Leslie, P. (2009). Consumer boycotts: The impact of the iraq war on french wine sales in the us. *QME*, 7(1):37–67.

Hainmueller, J., Hiscox, M. J., and Sequeira, S. (2015). Consumer demand for fair trade: Evidence from a multistore field experiment. *Review of Economics and Statistics*, 97(2):242–256.

Hart, O., Thesmar, D., and Zingales, L. (2024). Private sanctions. *Economic Policy*, 39(117):203–268.

Hiscox, M. J. and Smyth, N. (2011). Is there consumer demand for fair labor standards? evidence from a field experiment. *Evidence from a Field Experiment (April 22, 2011)*.

Kaufmann, M., Andre, P., and Kőszegi, B. (2024). Understanding markets with socially responsible consumers. *The Quarterly Journal of Economics*, 139(3):1989–2035.

King, B. G. (2008). A political mediation model of corporate response to social movement activism. *Administrative science quarterly*, 53(3):395–421.

Kitzmueller, M. and Shimshack, J. (2012). Economic perspectives on corporate social responsibility. *Journal of economic literature*, 50(1):51–84.

Klein, J. G., Smith, N. C., and John, A. (2004). Why we boycott: Consumer motivations for boycott participation. *Journal of marketing*, 68(3):92–109.

Pandya, S. S. and Venkatesan, R. (2014). French roast: Consumer responses to international conflict-evidence from supermarket scanner data. *Review of Economics and Statistics, Forthcoming*.

Pigors, M. and Rockenbach, B. (2016). Consumer social responsibility. *Management Science*, 62(11):3123–3137.

Sonnenfeld, J., Tian, S., Zaslavsky, S., Bhansali, Y., and Vakil, R. (2022). It pays for companies to leave russia. *Available at SSRN*, 4112885.

Appendix of Proofs

Proof of Lemma 1. Fix $p \in (0, \bar{p})$. Since $p < \bar{p} = U'(0^+)$, ethical demand $x_1(p)$ is strictly positive and interior, solving $U'(x_1) = p$.

Part (1). If $\gamma = 0$ the two first-order conditions in (2) coincide, so $x_0(p, 0) = x_1(p)$. Suppose $\gamma > 0$. If unethical demand is interior, it solves $U'(x_0) - \gamma C'(x_0) = p$, that is, $U'(x_0) = p + \gamma C'(x_0) > p = U'(x_1)$, where the strict inequality uses $C' > 0$. Since $U'' < 0$, U' is strictly decreasing, so $x_0(p, \gamma) < x_1(p)$. If instead unethical demand is at the corner $x_0(p, \gamma) = 0$, then $x_0(p, \gamma) = 0 < x_1(p)$ trivially. In all cases $x_0(p, \gamma) \leq x_1(p)$, with equality only if $\gamma = 0$.

Part (2). Ethical demand solves $U'(x_1) = p$, which does not involve γ, so $\partial x_1/\partial\gamma = 0$. For interior unethical demand, differentiate the identity $U'(x_0(p,\gamma)) - \gamma C'(x_0(p,\gamma)) = p$ with respect to γ:

$$[U''(x_0) - \gamma C''(x_0)]\,\frac{\partial x_0}{\partial \gamma} - C'(x_0) = 0 \quad \Longrightarrow \quad \frac{\partial x_0}{\partial \gamma} = \frac{C'(x_0)}{U''(x_0) - \gamma C''(x_0)}.$$

The numerator is positive by Assumption 2, and the denominator is negative because $U'' < 0$, $C'' > 0$, and $\gamma \geq 0$. Hence $\partial x_0/\partial\gamma < 0$. □

Proof of Lemma 2. Ethical profit $\pi_1 = \max_p (p - c_1)x_1(p)$ does not depend on γ, because x_1 does not; hence $\pi_1'(\gamma) = 0$.

For unethical profit, let $p_0(\gamma)$ denote the optimal price, which is unique and interior by Assumption 3, with $p_0 > c_0$ since the firm can guarantee positive profit. Applying the envelope theorem to $\pi_0(\gamma) = \max_p (p - c_0)x_0(p,\gamma)$,

$$\pi_0'(\gamma) = (p_0 - c_0)\,\frac{\partial x_0}{\partial \gamma}\bigg|_{p=p_0}.$$

The markup $p_0 - c_0 > 0$, and $\partial x_0/\partial\gamma < 0$ by Lemma 1(2), so $\pi_0'(\gamma) < 0$. Therefore $\Delta\pi'(\gamma) = \pi_1'(\gamma) - \pi_0'(\gamma) = -\pi_0'(\gamma) > 0$. □

Proof of Lemma 3. Ethical welfare $W_1 = U(x_1) - c_1 x_1$ does not depend on γ, so $W_1'(\gamma) = 0$.

We first show that the equilibrium quantity under unethical production is strictly decreasing in γ. Write the firm's problem in quantity space. The inverse demand under unethical production is $P(x,\gamma) = U'(x) - \gamma C'(x)$, obtained from the consumer's first-order condition, so the firm chooses x to maximize

$$\Pi(x,\gamma) = (U'(x) - \gamma C'(x) - c_0)\,x.$$

The first-order condition is $\Pi_x = U'(x) - \gamma C'(x) - c_0 + x\,(U''(x) - \gamma C''(x)) = 0$, and Assumption 3 delivers a unique maximizer $x_0(\gamma)$ with $\Pi_{xx} < 0$ there. The cross-partial is

$$\Pi_{x\gamma} = -C'(x) - xC''(x) < 0,$$

since $C' > 0$, $C'' > 0$, and $x > 0$. By the implicit function theorem,

$$\frac{dx_0}{d\gamma} = -\frac{\Pi_{x\gamma}}{\Pi_{xx}} < 0,$$

since $-\Pi_{x\gamma} > 0$ and $\Pi_{xx} < 0$. So equilibrium quantity strictly decreases in γ.

Now differentiate unethical welfare $W_0(\gamma) = U(x_0) - \gamma C(x_0) - c_0 x_0$ along the equilibrium path:

$$\frac{dW_0}{d\gamma} = [U'(x_0) - \gamma C'(x_0) - c_0] \frac{dx_0}{d\gamma} - C(x_0).$$

By the consumer's first-order condition, $U'(x_0) - \gamma C'(x_0) = p_0$, so the bracket equals the markup $p_0 - c_0 > 0$. Wherever unethical demand is positive, $dx_0/d\gamma < 0$ makes the first term negative and $-C(x_0) < 0$ by Assumption 2. Hence $W_0'(\gamma) < 0$. Consequently $\Delta W'(\gamma) = W_1'(\gamma) - W_0'(\gamma) = -W_0'(\gamma) > 0$. □

Proof of Proposition 1. The firm's unethical objective is $\Pi_0(p) = (p - c_0)x_0(p, \gamma)$, and its maximized value $\pi_0(\gamma)$ depends on γ only through the demand $x_0(p, \gamma)$. By the consumer's first-order condition $U'(x_0) - \gamma C'(x_0) = p$, that demand depends on γ only through the marginal moral cost C' and its derivatives; the level $C(x_0)$ never enters. Ethical profit is independent of γ. Hence $\Delta\pi$ depends on the moral preference only through the demand response it induces. By contrast, unethical welfare $W_0 = U(x_0) - \gamma C(x_0) - c_0 x_0$ contains the term $\gamma C(x_0)$ explicitly, so ΔW depends on the level C in addition to the demand response. The two comparisons therefore respond to different features of the same moral-cost function. □

Proof of Lemma 4. By Lemma 2, $\Delta\pi$ is continuous and strictly increasing, with $\Delta\pi(0) < 0$ and $\lim_{\gamma\to\infty} \Delta\pi(\gamma) = \pi_1 > 0$. The intermediate value theorem yields a value γ_π with $\Delta\pi(\gamma_\pi) = 0$; strict monotonicity makes it unique, and it is finite because the sign change occurs at a finite argument. By Lemma 3, ΔW is continuous and strictly increasing, with $\Delta W(0) < 0$. As $\gamma \to \infty$, unethical demand vanishes and $W_0 \to -\gamma C(0) < 0$, so ΔW eventually exceeds $W_1 > 0$; the same argument gives a unique, finite γ_W with $\Delta W(\gamma_W) = 0$. □

Proof of Proposition 2. By Lemma 2, $\Delta\pi(\gamma) \geq 0$ if and only if $\gamma \geq \gamma_\pi$, so the firm produces ethically exactly when $\gamma \geq \gamma_\pi$. By Lemma 3, $\Delta W(\gamma) \geq 0$ if and only if $\gamma \geq \gamma_W$, so ethical production is efficient exactly when $\gamma \geq \gamma_W$. The two decisions differ precisely when exactly one of the inequalities $\gamma \geq \gamma_\pi$ and $\gamma \geq \gamma_W$ holds, i.e., when γ lies strictly between the two thresholds. If $\gamma_W < \gamma_\pi$, then for $\gamma \in (\gamma_W, \gamma_\pi)$ ethical production is efficient while the firm produces unethically: the firm switches too late. If $\gamma_\pi < \gamma_W$, then for $\gamma \in (\gamma_\pi, \gamma_W)$ the firm produces ethically while unethical production is efficient: the firm switches too early. For γ outside the closed interval between the thresholds, both inequalities agree and the choices coincide. □

Proof of Theorem 1. Use the additive level of the moral-cost function as one coordinate of the primitive vector. Thus write $\vartheta = (a, \theta)$, where θ collects all remaining primitives and a change in a replaces the moral-cost function by

$$C_a(x) = C(x) + a.$$

Because $C(0) > 0$, sufficiently small changes in a remain admissible. Moreover,

$$C'_a(x) = C'(x), \qquad C''_a(x) = C''(x).$$

An additive change in a therefore leaves both demand curves, equilibrium prices and quantities, profits, and the private threshold $\gamma_\pi(\theta)$ unchanged.

Welfare does change. Under unethical production,

$$W_{0,a}(\gamma; \theta) = W_{0,0}(\gamma; \theta) - \gamma a,$$

while ethical welfare is unchanged. Hence

$$\Delta W_a(\gamma; \theta) = \Delta W_0(\gamma; \theta) + \gamma a.$$

Since γ_W is the unique zero of ΔW_a, threshold alignment is equivalent to

$$0 = \Delta W_a(\gamma_\pi(\theta); \theta) = \Delta W_0(\gamma_\pi(\theta); \theta) + \gamma_\pi(\theta) a.$$

By Lemma 4, $\gamma_\pi(\theta) > 0$. Therefore, conditional on θ, there is at most one additive level that produces alignment:

$$a^*(\theta) = -\frac{\Delta W_0(\gamma_\pi(\theta); \theta)}{\gamma_\pi(\theta)}.$$

The alignment set is thus contained in the graph $\{(a, \theta) : a = a^*(\theta)\}$, a codimension-one subset of the full parameter space. Equivalently, every section of the alignment set obtained by fixing θ contains at most one point in the a-direction. By Fubini's theorem, this set has Lebesgue measure zero. Any probability distribution absolutely continuous with respect to Lebesgue measure consequently assigns it probability zero, so

$$\mathbb{P}(\gamma_\pi = \gamma_W) = 0.$$

On the complementary event, Proposition 2 gives a non-empty open interval of moral concern on which the firm's regime choice is inefficient. □

Proof of Proposition 4. Ethical demand, and hence the ethical equilibrium, does not involve γ, so $\pi_1'(\gamma) = 0$ and $W_1'(\gamma) = 0$.

Since $P_X = U'' - \gamma C'' < 0$, Assumption 4(ii) gives, for every $N \geq 1$,

$$(N+1)P_X + XP_{XX} = \big(2P_X + XP_{XX}\big) + (N-1)P_X < 0. \tag{A.1}$$

Differentiating the equilibrium condition (14) with respect to γ, using $P_\gamma = -C'(X)$ and $P_{X\gamma} = -C''(X)$,

$$\frac{dX_0}{d\gamma} \;=\; \frac{C'(X_0) + \frac{X_0}{N}C''(X_0)}{\frac{N+1}{N}P_X + \frac{X_0}{N}P_{XX}} \;<\; 0,$$

the numerator positive by Assumption 2 and the denominator negative by (A.1).

For welfare, differentiate $W_0(\gamma) = U(X_0) - \gamma C(X_0) - c_0 X_0$ along the equilibrium path and use the consumer's first-order condition $U'(X_0) - \gamma C'(X_0) = p_0$:

$$\frac{dW_0}{d\gamma} = (p_0 - c_0)\frac{dX_0}{d\gamma} - C(X_0) < 0,$$

since the markup is positive for every finite N, $dX_0/d\gamma < 0$, and $C(X_0) > 0$. Hence $\Delta W'(\gamma) > 0$.

For profit, write firm i's payoff as $\pi_i = (P(X,\gamma) - c_0)q_i$ with $X = q_i + Q_{-i}$. At the symmetric equilibrium $q_i = X_0/N$ is optimal given Q_{-i}, so by the envelope theorem only the direct effect of γ and the induced change in rivals' output matter:

$$\frac{d\pi_0}{d\gamma} = q\,P_\gamma + q\,P_X \cdot \frac{N-1}{N}\frac{dX_0}{d\gamma} = \frac{X_0}{N}\Big[-C'(X_0) + \frac{N-1}{N}\,P_X\,\frac{dX_0}{d\gamma}\Big].$$

The first term inside the bracket is negative. The second is positive, since $P_X < 0$ and $dX_0/d\gamma < 0$, and it is strictly positive whenever $N \geq 2$. Assumption 4 places no bound on C'', and the second term is increasing in $C''(X_0)$ through $dX_0/d\gamma$; the bracket may therefore take either sign, and $\pi_0'(\gamma)$ is of ambiguous sign for every $N \geq 2$. At $N = 1$ the second term vanishes and $\pi_0'(\gamma) < 0$, recovering Lemma 2. □

Proof of Proposition 5. *Endpoints.* At $\gamma = 0$ the two regimes share the demand curve and differ only in marginal cost. For a symmetric Cournot industry facing a common cost c, the envelope theorem gives

$$\frac{d\pi}{dc} = q\Big[-1 + \frac{(N-1)P_X}{(N+1)P_X + XP_{XX}}\Big] < 0,$$

where the inequality holds because Assumption 4(ii) is equivalent to $(N+1)P_X + XP_{XX} <$

$(N-1)P_X$, which with (A.1) places the ratio in $[0,1)$. Similarly $dW/dc = (P-c)\frac{dX}{dc} - X < 0$, since $dX/dc = \left[\frac{N+1}{N}P_X + \frac{X}{N}P_{XX}\right]^{-1} < 0$ by (A.1). Hence $c_0 < c_1$ implies $\Delta\pi(0) < 0$ and $\Delta W(0) < 0$. As $\gamma \to \infty$, unethical output vanishes, so $\pi_0 \to 0$ and $W_0 \to -\gamma C(0) < 0$, giving $\Delta\pi \to \pi_1 > 0$ and ΔW eventually exceeding $W_1 > 0$.

Thresholds. $\Delta\pi$ is continuous, negative at 0, and eventually positive, so $Z := \{\gamma \geq 0 : \Delta\pi(\gamma) = 0\}$ is non-empty by the intermediate value theorem. Z is closed by continuity and bounded, since $\Delta\pi$ is bounded away from zero for large γ; hence Z is compact and $\underline{\gamma}_\pi = \min Z$ and $\overline{\gamma}_\pi = \max Z$ are attained and finite. By minimality $\Delta\pi < 0$ on $[0, \underline{\gamma}_\pi)$, and by maximality $\Delta\pi > 0$ on $(\overline{\gamma}_\pi, \infty)$. By Proposition 4, ΔW is continuous and strictly increasing with $\Delta W(0) < 0$ and ΔW eventually positive, so γ_W is unique and finite, with $\Delta W < 0$ below it and $\Delta W > 0$ above.

Regions. The industry produces ethically exactly when $\Delta\pi \geq 0$; ethical production is efficient exactly when $\Delta W \geq 0$. The choice is efficient at γ if and only if the two comparisons agree in sign. For $\gamma < \min\{\gamma_W, \underline{\gamma}_\pi\}$ both are negative and for $\gamma > \max\{\gamma_W, \overline{\gamma}_\pi\}$ both are positive, giving claim 1. If $\gamma_W < \underline{\gamma}_\pi$, then on $(\gamma_W, \underline{\gamma}_\pi)$ we have $\Delta W > 0 > \Delta\pi$, so ethical production is efficient while the industry produces unethically, giving claim 2; the symmetric argument on $(\overline{\gamma}_\pi, \gamma_W)$ gives claim 3. Claim 4 is the definition of sign disagreement, which by the previous sentences can occur only inside $(\underline{\gamma}_\pi, \overline{\gamma}_\pi)$ in that case.

Efficiency everywhere. The choice is efficient at every γ if and only if $\{\gamma : \Delta\pi(\gamma) \geq 0\} = \{\gamma : \Delta W(\gamma) \geq 0\} = [\gamma_W, \infty)$. The right-hand equality is monotonicity of ΔW. The left-hand equality forces $\Delta\pi < 0$ on $[0, \gamma_W)$ and $\Delta\pi \geq 0$ on $[\gamma_W, \infty)$, so $Z \subseteq [\gamma_W, \infty)$ with $\gamma_W = \min Z$, and no zero of $\Delta\pi$ can be followed by a return to negative values; hence $\underline{\gamma}_\pi = \overline{\gamma}_\pi = \gamma_W$. The converse is immediate. □

Proof of Proposition 6. As in the proof of Theorem 1, write the primitive vector as $\vartheta = (a, \theta)$, where changing a replaces C by $C_a = C + a$ and θ collects the remaining primitives. This perturbation leaves inverse demand, equilibrium output, per-firm profit, $\Delta\pi$, and therefore the entire zero set of $\Delta\pi$, including $\underline{\gamma}_\pi(\theta)$, unchanged.

It shifts unethical welfare by $-\gamma a$, so

$$\Delta W_a(\gamma; \theta) = \Delta W_0(\gamma; \theta) + \gamma a.$$

A necessary condition for efficiency at every level of moral concern is $\underline{\gamma}_\pi = \gamma_W$, which is equivalent to

$$0 = \Delta W_a\Big(\underline{\gamma}_\pi(\theta); \theta\Big) = \Delta W_0\Big(\underline{\gamma}_\pi(\theta); \theta\Big) + \underline{\gamma}_\pi(\theta) a.$$

Since $\underline{\gamma}_\pi(\theta) > 0$, for each fixed θ this equation has at most one solution in a. The coincidence

set is therefore contained in a codimension-one graph in the full parameter space and has Lebesgue measure zero by Fubini's theorem. Any absolutely continuous distribution over the primitives assigns it probability zero.

On the complementary event $\underline{\gamma}_{\pi} \neq \gamma_W$, Proposition 5 implies that the industry's regime choice is inefficient on a non-empty open set of moral concern. $\square$