Joint Discussion Paper Series in Economics

by the Universities of
Aachen • Gießen • Göttingen
Kassel • Marburg • Siegen

ISSN 1867-3678

No. 10-2026

Gerrit Gräper and Fabian Mankat

When green technology adoption backfires: A theoretical analysis with heterogeneous internalization of environmental harm

Coordination: Bernd Hayo
Philipps-University Marburg
School of Business and Economics
Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

When green technology adoption backfires: A theoretical analysis with heterogeneous internalization of environmental harm

Gerrit Gräper*, Fabian Mankat†

July 21, 2026

Abstract

Policies that promote green technology adoption are a central tool for reducing emissions, but efficiency improvements can induce rebound through higher post adoption usage. When this response is strong enough, total emissions increase, an outcome known as the backfire effect. We develop a theoretical model in which individuals differ in how strongly they internalize environmental harm, and we characterize adoption and usage as functions of this internalization, demand elasticities, and material incentives. The model delivers three policy relevant insights. First, we derive conditions under which the backfire effect arises and show that it occurs only for sufficiently low internalization individuals, who increase usage strongly when per unit costs fall. Second, increasing incentives to adopt greener technologies can generate aggregate backfire through selection if they induce relatively many low internalization types to adopt. Third, such adoption incentives need not be monotone in internalization, because low types value cost savings while high types value emission reductions, so intermediate types may have the weakest private incentive to adopt. This creates a targeting dilemma for policymakers: inducing adoption among emission reducing non-adopters without drawing in marginal adopters most prone to backfire.

1 Introduction

Reducing emissions and mitigating environmental harm are among the most pressing challenges facing modern societies. Addressing these challenges requires individuals to shift toward greener technologies that produce lower emissions per unit of use, such as electric vehicles, heat pumps, or energy-efficient housing renovations. A growing body of evidence shows that efficiency improvements can give rise to *rebound effects*, whereby the adoption of greener technologies leads to more intensive use of the associated service, thereby attenuating the expected emissions reductions (Gillingham et al., 2016; Sorrell et al., 2009). In some cases, this behavioral response has been documented to be sufficiently strong that total emissions even increased, a phenomenon commonly referred to as the *backfire effect* (Alcott, 2005; Jevons, 1865). Hence, while rebound effects primarily attenuate the effectiveness of policies that promote green technology adoption, they become a central policy concern when they escalate into backfire, as such policies may then exhibit adverse effects and increase total emissions.

This paper argues that the occurrence of backfire depends fundamentally on the extent to which individuals internalize environmental harm. While existing theoretical models typically explain rebound effects through prices and demand elasticities (e.g., Binswanger, 2001; Borenstein, 2015; Greening et al., 2000), they largely abstract from heterogeneity in how individuals evaluate emissions.[1] We develop a theoretical model that places this heterogeneity at the center and characterize how green technology adoption affects both individual and aggregate emissions. Specifically, we solve for optimal technology choice and usage as functions of internalization, per-unit prices, adoption costs, and demand elasticities, and derive conditions under which increased adoption reduces emissions, as well as conditions under which it leads to backfire.

Our main insights are threefold. First, we show that backfire depends critically on the degree to which adopters internalize environmental harm. Individuals who strongly internalize emissions do not increase their

*University of Kassel, Nora-Platiel-Straße 4, 34127 Kassel (Germany), E-mail: gerrit.graeper@uni-kassel.de

†Corresponding author, University of Kassel, Nora-Platiel-Straße 4, 34127 Kassel (Germany), E-mail: fabian.mankat@uni-kassel.de

[1] On heterogeneous and persistent internalization of environmental harm, see the empirical evidence listed in Dannenberg et al. (2024) and the theoretical frameworks in Mankat (2024, 2026).

use much upon adopting greener technologies, which translates into significant emission reductions. By contrast, individuals who internalize emissions only weakly respond to lower per-unit prices by increasing usage enough to offset part (and potentially most) of the per-unit emission advantage. Backfire is therefore not just dependent on prices and elasticities, but also on how they interact with internalization. We derive formal conditions under which backfire is likely to occur, providing policymakers with insights when this poses a viable threat.

Second, the model illustrates why strengthening external incentives to adopt greener technologies, such as adoption subsidies or social pressure, can lead to aggregate backfire effects, even when these incentives are designed to increase take-up rather than subsidize per-unit usage. The key mechanism is selection: external incentives affect not only how many individuals adopt the green technology, but also who adopts it. As incentives increase, the pool of adopters expands toward individuals for whom adoption is driven primarily by cost savings rather than by concerns about emissions, and who therefore exhibit particularly strong post-adoption usage responses. Consequently, aggregate emissions may increase. Whether such outcomes arise depends critically on the distribution of internalization parameters in the population. The policy implication is that external incentives should be applied with care and, where necessary, complemented by policies that strengthen the internalization of environmental harm, in order to ensure that increased adoption translates into meaningful emission reductions.

Third, we show that adoption of green technologies is not necessarily monotone with respect to the internalization of environmental harm. The reason is that low-internalization individuals value cost savings most, while high-internalization individuals value emission reductions most. Hence, the private incentive to adopt greener technologies can be weakest for intermediate types. This non-monotonicity poses a targeting dilemma for policymakers: if increasing external incentives increases aggregate emissions, there are non-adopters who would reduce emissions if they adopted the green technology. The policy problem that arises is therefore how to induce adoption among the types for whom adoption is emission-reducing, without expanding adoption among the types for whom adoption mainly increases usage.

This analysis contributes to the broader theoretical literature on rebound and backfire effects. Most prior research examines rebound effects from a macroeconomic perspective, identifying conditions under which efficiency improvements generate rebound at the aggregate level through economy-wide adjustments (see, e.g., Alcott, 2005; Christou et al., 2025; Grepperud & Rasmussen, 2004; Saunders, 1992, 2000; Tan & Walheer, 2025; Turner, 2009; Wei, 2007). In contrast, the present study focuses on the demand-side literature, which investigates how individual decisions and behavioral responses influence emissions. This body of work highlights the role of prices, income effects, and demand elasticities in shaping post-adoption behavior (Gillingham et al., 2016). The analysis specifically addresses cases where these behavioral responses are sufficiently strong to negate the intended efficiency gains, resulting in backfire effects. Borenstein (2015) offers a microeconomic framework in which substitution and income effects following efficiency improvements can, in principle, offset emission savings. Related studies examine similar demand-side mechanisms at the household level, including conceptual and partial-equilibrium analyses of energy efficiency and usage responses (e.g., Binswanger, 2001; Greening et al., 2000). In comparison to these analyses, this study incorporates heterogeneity in the internalization of environmental harm as a central mechanism influencing both adoption and post-adoption usage. Furthermore, adoption of green technologies, and thus efficiency gains, follows an endogenous decision and is responsive to external incentives, which allows for the characterization of how policy-induced changes affect aggregate outcomes.

The remainder of the paper presents the model and formal results, followed by a discussion of the findings and their implications.

2 Formal analysis

2.1 Theoretical framework

We consider a continuum of individuals $i \in I = [0, 1]$. All individuals take two decisions. First, each individual i chooses a technology $t_i \in \{g, b\}$, where g and b stand for green and brown, respectively. Second, conditional on the chosen technology, the individual chooses a level of use $x_i \geq 0$. Technologies differ in their per-unit price p_t, emissions per unit e_t, and fixed adoption cost k_t. We assume $e_g < e_b$, $p_g < p_b$, and $k_b < k_g$: the green technology is both cleaner and cheaper per-unit, but requires a higher fixed cost of adoption.

Each individual derives benefits from use x_i according to a CRRA utility function:

$$b(x_i) = \frac{x_i^{1-\rho} - 1}{1-\rho},$$

where $\rho > 1$ implies a low demand elasticity, given by $1/\rho$, consistent with empirical evidence on the rigidity of energy and transport demand (Labandeira et al., 2017; Mubiinzi et al., 2024).

Moreover, each individual incurs material costs consisting of a fixed technology adoption cost k_{t_i} and total variable expenditure $p_{t_i} x_i$.

In addition to material costs, individuals experience personal sanctions from emitting. These personal sanctions are proportional to the individual's emissions $e_{t_i} x_i$ and scale with their internalization parameter $m_i \geq 0$. The intuitive idea is that m_i measures the degree to which an individual i internalizes environmental harm. Hence, personal sanctions equal $m_i e_{t_i} x_i$. We assume that m_i is distributed according to cumulative distribution function $F(\cdot)$ with corresponding probability density function $f(\cdot)$.

Combining the above elements, the utility of an individual of type m_i who uses x_i units of technology t_i is given by

$$U(x_i, t_i, m_i, \hat{x}_{t_i}) = b(x_i) - m_i e_{t_i} x_i - p_{t_i} x_i - k_{t_i}.$$

We assume that material costs and the internalized harm from emission enter utility linearly, which allows for transparent comparative statics. Relaxing this assumption would not qualitatively affect the main results.

2.2 Optimal use for given technologies

We begin by analyzing individual i's optimal usage x_i for a given technology t. Let $x^*(m_i, t)$ denote optimal consumption as a function of the internalization parameter m_i and the chosen technology. Solving the first-order condition of the utility function yields

$$x^*(m_i, t) = (p_t + m_i e_t)^{-1/\rho}.$$

Optimal usage and total emissions $e_t x^*(m_i, t)$ decrease in the per-unit price p_t, per-unit emissions e_t, and internalization parameter m_i.

Since the green technology has lower per-unit prices and emissions, $p_g < p_b$ and $e_g < e_b$, it must hold that each individual has increased use under the green relative to the brown technology, $x^*(m_i, g) > x^*(m_i, b)$. Hence, each individual experiences a rebound effect in usage. This raises the possibility that higher usage may outweigh lower per-unit emissions, leading to higher total emissions under the green technology. The following proposition captures our formal results concerning this backfire effect.

Proposition 2.1. *Let* $\bar{m} = \frac{p_b e_b^{-\rho} - p_g e_g^{-\rho}}{e_g^{1-\rho} - e_b^{1-\rho}}$.

1.

$$m_i < \bar{m} \quad \Leftrightarrow \quad e_g\, x^*(m_i, g) > e_b\, x^*(m_i, b)$$

2.

$$\bar{m} > 0 \quad \Leftrightarrow \quad \frac{e_g}{e_b} > \left(\frac{p_g}{p_b}\right)^{\frac{1}{\rho}}$$

Proof: See Appendix A.

Condition 1 shows that total emissions are higher under the green technology if and only if the individual's internalization parameter lies below the threshold $\bar{m}$. Condition 2 characterizes when such a threshold exists: backfire effects arise only if the relative reduction in per-unit prices p_g/p_b is sufficiently large compared to the relative per-unit emissions e_g/e_b, and the elasticity $1/\rho$ is not too low. Intuitively, for individuals with weak internalization, the lower per-unit price of the green technology induces a strong usage response. When this response is large enough, the increase in usage dominates the per-unit emission reduction, leading to higher total emissions.

Two key implications follow. First, adoption of green technologies can increase emissions for a subset of individuals, depending on the degree of their internalization. Second, lowering per-unit prices of green technologies facilitates backfire effects by expanding the range of individuals for whom adoption increases usage sufficiently to raise total emissions.

2.3 Technology choice

We now turn to individuals' technology adoption decisions. Because technology choice precedes usage, individuals anticipate their usage, and technology choices are thus made under perfect foresight regarding subsequent behavior. We are interested in the relative attractiveness of the green and brown technologies. To this end, we write the difference in utilities as

$$U(x^*(m_i, g),\, g,\, m_i) - U(x^*(m_i, b),\, b,\, m_i) = \frac{\rho}{1-\rho}\Big[(p_g + m_i e_g)^{\frac{\rho-1}{\rho}} - (p_b + m_i e_b)^{\frac{\rho-1}{\rho}}\Big] - \Big[k_g - k_b\Big].$$

The above difference consists of two parts. The first one is

$$\Delta^{use}(m_i) := \frac{\rho}{1-\rho}\Big[(p_g + m_i e_g)^{\frac{\rho-1}{\rho}} - (p_b + m_i e_b)^{\frac{\rho-1}{\rho}}\Big]$$

which captures the variable utility difference from green and brown technology usage, including the per-unit prices and internalization parameter associated with usage. Note that this difference depends on the per-unit prices (p_g, p_b), the internalization parameter m_i, per-unit emissions (e_g, e_b), and the elasticity ρ. These variables affect the optimal quantities as well as the costs associated with adjusting them. Thus, $\Delta^{use}(m_i)$ excludes the fixed costs associated with adopting the technologies.

The second part is

$$\Delta^{fix} := k_g - k_b,$$

which captures the fixed material and social costs of adopting the technologies. In particular, it depends on the material costs of adoption as well as the social sanctions associated with each technology.

Any individual i strictly prefers to adopt the green technology if and only if

$$\Delta^{use}(m_i) > \Delta^{fix}.$$

We employ the following assumption, which ensures that there is an internal division of individuals into both technologies.[2]

Assumption 2.1. $\min_{\{m|f(m)>0\}}[\Delta^{use}(m)] < \Delta^{fix} < \max_{\{m|f(m)>0\}}[\Delta^{use}(m)]$

Note that Δ^{fix} is constant in m_i. Hence, to understand when $\Delta^{use}(m_i) > \Delta^{fix}$ holds, we need to look at the shape of $\Delta^{use}(m_i)$.

Lemma 2.1. *Let* $\bar{m} = \frac{p_b\, e_b^{-\rho} - p_g\, e_g^{-\rho}}{e_g^{1-\rho} - e_b^{1-\rho}}$.

1.
$$\frac{\partial \Delta^{use}(m_i)}{\partial m_i} < 0 \Leftrightarrow m_i < \bar{m}$$

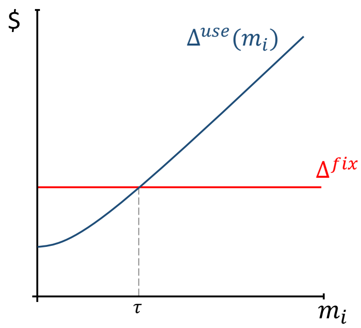
2.
$$\frac{\partial \Delta^{use}(m_i)}{\partial m_i} = 0 \Leftrightarrow m_i = \bar{m}$$

3.
$$\frac{\partial \Delta^{use}(m_i)}{\partial m_i} > 0 \Leftrightarrow m_i > \bar{m}$$

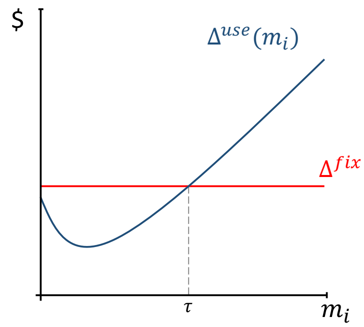
4.
$$\lim_{m_i \to \infty} \Delta^{use}(m_i) = \infty$$

Proof: See Appendix A.

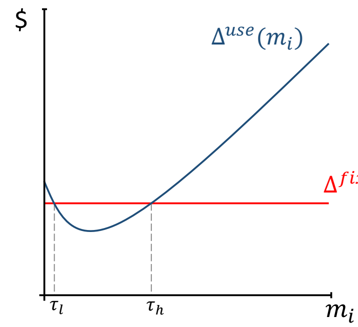
[2] Note that the results do not change when dropping this, but the writing becomes a bit more messy.



(a) $\bar{m} \leq 0$



(b) $\bar{m} > 0$ and $\Delta^{use}(0) < \Delta^{fix}$



(c) $\bar{m} > 0$ and $\Delta^{use}(0) > \Delta^{fix}$

Figure 1: Utility from green technology adoption dependent on the internalization parameter m_i.

The above lemma states that the relative benefit $\Delta^{use}(m_i)$ is decreasing in m_i for $m_i < \bar{m}$, and increasing in m_i for $m_i > \bar{m}$. Note that this threshold $\bar{m}$ is the same as in Proposition 2.1. Thus, individuals with internalization parameter $m_i < \bar{m}$ are exactly those who emit more when using the green than when using the brown technology. Consequently, the benefit from adopting the green technology is possibly non-monotone in the internalization parameter m_i. Moreover, Condition 4 implies that as m_i tends to infinity, so does the relative use benefit from the green technology.

The intuition behind the result is as follows. First, consider an individual with internalization parameter $m_i < \bar{m}$. Such an individual experiences a backfire effect upon adopting the green technology: due to the lower per-unit price, they emit more than they would when using the brown technology. Hence, their main incentive to adopt the green technology is not emission reduction, but the decrease in the per-unit price. A marginal increase in the internalization parameter m_i decreases their consumption level x_g^*, and, consequently, the cheaper per-unit price has a smaller effect on total utility.

Second, consider an individual with internalization parameter $m_i > \bar{m}$. Such an individual always emits less under the green than under the brown technology. The main incentive for adopting the green technology is therefore to reduce total emissions. Hence, the stronger the internalization parameter, the larger the utility gain from emission reduction.

Figure 1 depicts three possible cases for how $\Delta^{use}(m_i)$ and Δ^{fix} may look. The first example corresponds to $\bar{m} \leq 0$: all individuals always emit less using the green technology. In that case, there is a unique threshold τ so that all individuals with an internalization parameter m_i below that threshold prefer not to adopt the green technology. Figure 1b depicts the case where $\Delta^{use}(m_i)$ is non-monotone, but the fixed adoption cost difference is so large that only individuals with a high internalization parameter adopt the green technology. Lastly, Figure 1c depicts the case where $\Delta^{use}(m_i)$ is non-monotone, and Δ^{fix} is relatively low. In that case, individuals with high internalization parameters, e.g., $m_i > \tau^h$, and individuals with low internalization parameters, e.g., $m_i < \tau^l$, both prefer to adopt the green technology. Those with intermediate internalization parameters prefer to stick to the brown technology.

From Figure 1b and Figure 1c it is apparent that the set of internalization parameters m_i for which individuals adopt the brown technology always corresponds to a connected interval. The following proposition captures this formally.

Proposition 2.2. *There is a connected and non-empty interval $M^b \in [0, \infty)$ s.t. any individual i weakly prefers to adopt the brown technology if and only if $m_i \in M^b$.*

Proof: See Appendix A.

2.4 External incentives and total emissions

In this section, we study how external incentives to adopt greener technologies affect total emissions. Such incentives can take the form of monetary subsidies for green technologies or social pressure and stigma associated with brown technologies. We deliberately focus on incentives that operate at the technology adoption stage.

The reason is twofold. First, in the context of monetary policy, subsidies aimed at reducing emissions are generally argued to primarily target the fixed costs of adoption rather than per-unit prices, since lower per-unit prices tend to increase usage and may therefore increase emissions, a mechanism that is also present in our framework. Second, social incentives such as stigma or reputational concerns are typically tied to technology adoption rather than usage, since the adopted technology is often publicly observable while individual usage levels are generally not. For example, others can easily observe which type of car an individual drives, but they cannot monitor how much the car is actually used.

For these reasons, we study the effect of changes in Δ^{fix}. Although formally introduced as the difference in monetary adoption costs, Δ^{fix} can be interpreted more broadly to include social stigma, reputational rewards, or other forms of external pressure that affect technology adoption without directly influencing usage behavior. An external incentive to adopt green technologies thus corresponds to a decrease in Δ^{fix}.[3]

Consider the set of internalization parameters m_i for which an individual weakly prefers to adopt the brown technology as a function of Δ^{fix}: $M^b(\Delta^{fix})$. The total emission can then be written as

$$E(\Delta^{fix}) = \int_0^{\min[M^b(\Delta^{fix})]} x^*(z,g)\, e_g f(z)dz + \int_{\min[M^b(\Delta^{fix})]}^{\max[M^b(\Delta^{fix})]} x^*(z,b)\, e_b f(z)dz + \int_{\max[M^b(\Delta^{fix})]}^{\infty} x^*(z,g)\, e_g f(z)dz$$

To study the effect on total emission $E(\cdot)$, we consider marginal changes in Δ^{fix}. First, we derive the result where there is a unique intersection so that all individuals with small internalization parameters adopt the brown and all with high adopt the green technology. This corresponds to the cases in Figure 1a and Figure 1b, and requires that Δ^{fix} is so high that individuals who do not care for emissions at all, $m_i = 0$, prefer not to adopt the green technology, $\Delta^{fix} > \Delta^{use}(0)$.

Lemma 2.2. *Suppose $\Delta^{fix} > \Delta^{use}(0)$.*

$$\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \geq 0.$$

Proof: See Appendix A.

The lemma states that in that case, a small decrease in Δ^{fix} leads to a decrease in total emission. The small decrease in Δ^{fix} corresponds to a downwards shift of Δ^{fix} in Figure 1a and Figure 1b, which induces individuals to adopt the green technology. These individuals, however, are only those who have an internalization parameter m_i above $\bar{m}$, and, hence, reduce their emission. Consequently, total emission must go down.

Next, consider Figure 1c, where $\bar{m} > 0$, so that some individuals exhibit a backfire effect when adopting the green technology. Moreover, since $\Delta^{fix} < \Delta^{use}(0)$, individuals with very low internalization weakly prefer the green technology.

External incentives that promote the adoption of greener technologies effectively decrease Δ^{fix} and induce two distinct groups of individuals to switch. First, individuals with relatively high internalization parameters m_i, for whom adopting the green technology reduces total emissions. Second, individuals with relatively low m_i, for whom adoption increases total emissions due to higher usage. The following proposition establishes that a marginal decrease in Δ^{fix} reduces (increases) aggregate emissions if the mass of newly adopting individuals who reduce emissions exceeds (falls short of) the mass of those who increase emissions.[4]

Lemma 2.3. *Suppose $\Delta^{fix} < \Delta^{use}(0)$.*

1.

$$\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \geq 0 \Leftrightarrow f(\max[M^b(\Delta^{fix})]) \geq f(\min[M^b(\Delta^{fix})]).$$

[3] In principle, the stigma associated with a technology could be defined endogenously, for example as a function of the average emissions of its adopters. Doing so would leave the main insights of this paper unchanged but would give rise to more complex technology adoption behavior due to interdependencies across individuals.

[4] Note that $E(\Delta^{fix})$ is continuous but not differentiable at $\Delta^{fix} = \Delta^{use}(0)$. The right-hand derivative at this point, i.e., $\lim_{\Delta^{fix} \downarrow \Delta^{use}(0)} \frac{E(\Delta^{fix}) - E(\Delta^{use}(0))}{\Delta^{fix} - \Delta^{use}(0)}$, is analogous to the derivative characterized in Lemma 2.2, while the left-hand derivative, i.e., $\lim_{\Delta^{fix} \uparrow \Delta^{use}(0)} \frac{E(\Delta^{fix}) - E(\Delta^{use}(0))}{\Delta^{fix} - \Delta^{use}(0)}$, is analogous to the derivative characterized in Lemma 2.3.

2.

$$\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \leq 0 \Leftrightarrow f(\max[M^b(\Delta^{fix})]) \leq f(\min[M^b(\Delta^{fix})]).$$

Proof: See Appendix A.

The proposition highlights that external adoption incentives may backfire: when the distribution of m_i places sufficient weight on low-internalization individuals, lowering Δ^{fix} can increase total emissions. Whether such policies are beneficial therefore depends critically on the distribution of internalization in the population.

Lemma 2.4. *If* $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \leq 0$*, then there is* $m_i \in M^b(\Delta^{fix})$ *s.t.* $e_g x^*(m_i, g) < e_b x^*(m_i, b)$*.*

Proof: See Appendix A.

Finally, Lemma 2.4 highlights a selection dilemma: When an increase in external incentives to adopt the green technology leads to an increase in total emissions, then there are still some types of individuals m_i who would decrease their emission with the green technology. However, these individuals cannot be targeted without also drawing in others, who exhibit a personal backfire effect and increase emission.

3 Discussion

This paper has shown how individual backfire effects depend on the degree to which an individual internalizes environmental harm. Individuals with sufficiently low internalization may emit more after adopting a green technology, because lower per unit prices induce a strong increase in usage when demand is sufficiently rigid and the per unit emission advantage of the green technology is not too large (Proposition 2.1). As a consequence, external incentives to adopt greener technologies can have an adverse effect on total emissions, because they affect not only how many individuals adopt but also who adopts (Lemma 2.3). Finally, adoption incentives are not necessarily monotone in internalization (Lemma 2.1), which implies a selection and targeting dilemma. When raising external incentives increases total emissions, there may still be non-adopters who would reduce their emissions if they adopted (Lemma 2.4), but inducing them to adopt risks drawing in marginal adopters who are most prone to backfire.

The results therefore, highlight a central policy issue. Strengthening social pressure or providing monetary subsidies may appear effective if one focuses on green technology adoption, but could turn out to be counterproductive from an emissions perspective. Our model implies that policies should therefore not focus exclusively on adoption, but also on fostering the internalization of environmental harm, for example, through cultivating persistent pro-environmental attitudes. This holds especially true if the distribution of preferences is concentrated at low internalization values. Importantly, this paper does not argue against promoting green technologies. Increasing the use of greener technologies is likely unavoidable for reducing emissions. Rather, the analysis flags a potential pitfall and highlights why adoption incentives may need to be complemented by policies that strengthen internalization.

Whereas rebound effects are empirically well documented, the empirical evidence on backfire effects appears more limited (Gillingham et al., 2016). This does not necessarily imply that backfire is not a relevant policy concern. Instead, it may reflect that the individuals for whom backfire would arise, namely those with very low internalization and strong post-adoption usage responses, have not yet adopted green technologies. As adoption expands and green technologies become cheaper to use, policy interventions may increasingly draw in such individuals, thereby raising the risk of aggregate backfire. Thus, the absence of widespread empirical evidence for backfire today does not preclude it from becoming a relevant policy concern as adoption expands to individuals with weaker environmental internalization.

To gauge whether the model's backfire prediction is possibly empirically relevant, it is useful to relate the key parameters to benchmark estimates from the applied literature. Labandeira et al. (2017) provide a meta-analysis of demand elasticities and estimate an average long-run elasticity for driving (gasoline demand) of about 0.773. da Costa et al. (2025) estimate average greenhouse gas emissions of 182.9 g CO_2-eq per km for electric vehicles, compared to 258.5 g CO_2-eq per km for gasoline vehicles, yielding a relative emission per km of $182.9/258.5 \approx 0.708$. For these values of $1/\rho = 0.773$ and $e_g/e_b = 0.708$, the condition of potential backfire in our model (i.e., $\bar{m} > 0$ in Proposition 2.1) holds whenever the relative price per use p_g/p_b is roughly below 0.64. Such relative per-use prices seem plausible based on currently available cost comparisons between electric and conventional vehicles. Hence, there are empirically plausible parameter regions in which the model predicts backfire for sufficiently low internalization types. While this is by no means an empirical validation of the model, it illustrates that backfire can be a viable policy concern, at least in the transport context. In other domains, such as residential heating, demand elasticities tend to be smaller (Labandeira et al., 2017), which, all else equal, makes backfire less likely. A more comprehensive assessment would require country-specific estimates of per-use cost differences (which depend on taxes, tariffs, and electricity prices) and a careful treatment of application-specific demand structures and usage margins.

Several aspects of the analysis deserve further attention. First, we did not consider endogenous social effects, in particular how social approval or stigma may evolve with who adopts green technology. Such feedback could amplify adoption dynamics. Second, we abstracted from fixed emissions associated with adoption itself, such as production and installation emissions. Third, we did not model system costs. Since many green technologies are electricity-based, large-scale adoption can put pressure on generation and the grid and may require grid expansion and regulatory adjustments (see also Gräper and Mankat, 2025). Finally, while we focus on heterogeneity in internalization, other dimensions of heterogeneity may matter as well, for example, income heterogeneity through substitution and income effects emphasized in the rebound literature. Overall, the model clarifies when adoption policies reduce emissions as intended and when they risk adverse aggregate outcomes through selection into adoption and backfire.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used *Claude AI* and *ChatGPT* to improve language and readability, help structure and refine arguments, support brainstorming, and obtain a preliminary impression of relevant literature. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Alcott, B. (2005). Jevons' paradox. *Ecological Economics*, *54*(1), 9–21.

Binswanger, M. (2001). Technological progress and sustainable development: What about the rebound effect? *Ecological Economics*, *36*(1), 119–132.

Borenstein, S. (2015). A microeconomic framework for evaluating energy efficiency rebound and some implications. *The Energy Journal*, *36*(1), 1–22.

Christou, T., Lecca, P., & Salotti, S. (2025). Regional rebound effects of energy efficiency improvements in a spatial general equilibrium framework. *Energy Economics*, 108645.

da Costa, V. B. F., Bitencourt, L., Dias, B. H., Soares, T., de Andrade, J. V. B., & Bonatto, B. D. (2025). Life cycle assessment comparison of electric and internal combustion vehicles: A review on the main challenges and opportunities. *Renewable and Sustainable Energy Reviews*, *208*, 114988.

Dannenberg, A., Gutsche, G., Batzke, M. C., Christens, S., Engler, D., Mankat, F., Möller, S., Weingärtner, E., Ernst, A., Lumkowsky, M., et al. (2024). The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*, *18*(1), 165–187.

Gillingham, K., Rapson, D., & Wagner, G. (2016). The rebound effect and energy efficiency policy. *Review of Environmental Economics and Policy*.

Gräper, G., & Mankat, F. (2025). Towards efficient pricing in monopolistic markets: A theoretical analysis of cost-based revenue caps employing a base-year approach. *Energy Economics*, *144*, 108309.

Greening, L. A., Greene, D. L., & Difiglio, C. (2000). Energy efficiency and consumption—the rebound effect—a survey. *Energy policy*, *28*(6-7), 389–401.

Grepperud, S., & Rasmussen, I. (2004). A general equilibrium assessment of rebound effects. *Energy Economics*, *26*(2), 261–282.

Jevons, W. S. (1865). *The coal question: An inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*. Macmillan.

Labandeira, X., Labeaga, J. M., & López-Otero, X. (2017). A meta-analysis on the price elasticity of energy demand. *Energy policy*, *102*, 549–568.

Mankat, F. (2024). Cooperation, norms, and gene-culture coevolution. *Games and Economic Behavior*, *147*, 242–267.
Mankat, F. (2026). Norms as a solution to the tragedy of the commons: A co-evolutionary model. *Journal of Economic Behavior & Organization*, *243*, 107433.
Mubiinzi, G., Senyonga, L., Kaawaase, T. K., Wasswa, F., Adaramola, M. S., & Nantongo, M. (2024). Income and price elasticities of household electricity demand: A comparative systematic review of aggregated and disaggregated data studies. *Energy Reports*, *12*, 4449–4465.
Saunders, H. D. (1992). The khazzoom-brookes postulate and neoclassical growth. *The Energy Journal*, *13*(4), 131–148.
Saunders, H. D. (2000). A view from the macro side: Rebound, backfire, and khazzoom–brookes. *Energy policy*, *28*(6-7), 439–449.
Sorrell, S., Dimitropoulos, J., & Sommerville, M. (2009). Empirical estimates of the direct rebound effect: A review. *Energy policy*, *37*(4), 1356–1371.
Tan, Y., & Walheer, B. (2025). On decomposing the energy rebound effect. *Energy Economics*, 108770.
Turner, K. (2009). Negative rebound and disinvestment effects in response to an improvement in energy efficiency in the uk economy. *Energy Economics*, *31*(5), 648–666.
Wei, T. (2007). Impact of energy efficiency gains on output and energy use with cobb–douglas production function. *Energy Policy*, *35*(4), 2023–2030.

A Proofs

Proof of Proposition 2.1:

Proof. Let $\bar{m} = \frac{p_b\, e_b^{-\rho} - p_g\, e_g^{-\rho}}{e_g^{1-\rho} - e_b^{1-\rho}}$. First, consider $e_g\, x^*(m_i, g) > e_b\, x^*(m_i, b)$. Substituting $x^*(m_i, g) = (p_g + m_i e_g)^{-1/\rho}$ and $x^*(m_i, b) = (p_b + m_i e_b)^{-1/\rho}$, and then rearranging yields $m_i < \frac{p_b\, e_b^{-\rho} - p_g\, e_g^{-\rho}}{e_g^{1-\rho} - e_b^{1-\rho}} = \bar{m}$, which proves the first condition. Next, consider $\bar{m} = \frac{p_b\, e_b^{-\rho} - p_g\, e_g^{-\rho}}{e_g^{1-\rho} - e_b^{1-\rho}} > 0$. Rearranging yields $\frac{e_g}{e_b} > \left(\frac{p_g}{p_b}\right)^{1/\rho}$, which in turn proves the second condition. □

Proof of Lemma 2.1:

Proof. Taking the FOC of $\Delta^{use}(m_i) := \frac{\rho}{1-\rho}\left[(p_g + m_i e_g)^{\frac{\rho-1}{\rho}} - (p_b + m_i e_b)^{\frac{\rho-1}{\rho}}\right]$ yields $\frac{\partial \Delta^{use}(m_i)}{\partial m_i} = e_g\,(p_g + m_i e_g)^{\frac{-1}{\rho}} - e_b\,(p_b + m_i e_b)^{\frac{-1}{\rho}} = e_g\, x^*(m_i, g) - e_b\, x^*(m_i, b)$. From Proposition 2.1 we know that (1) $m_i < \bar{m} \Leftrightarrow e_g\, x^*(m_i, g) > e_b\, x^*(m_i, b)$, which in turn implies $m_i < \bar{m} \Leftrightarrow \frac{\partial \Delta^{use}(m_i)}{\partial m_i} < 0$. In analog fashion we can show that (2) $m_i = \bar{m} \Leftrightarrow e_g\, x^*(m_i, g) = e_b\, x^*(m_i, b) \Leftrightarrow \frac{\partial \Delta^{use}(m_i)}{\partial m_i} = 0$ and (3) $m_i > \bar{m} \Leftrightarrow e_g\, x^*(m_i, g) < e_b\, x^*(m_i, b) \Leftrightarrow \frac{\partial \Delta^{use}(m_i)}{\partial m_i} > 0$. Lastly, we consider the fourth condition. $\Delta^{use}(m_i) = \frac{\rho}{1-\rho}\left[(p_g + m_i e_g)^{\frac{\rho-1}{\rho}} - (p_b + m_i e_b)^{\frac{\rho-1}{\rho}}\right] = \frac{\rho}{1-\rho} m_i^{\frac{\rho-1}{\rho}}\left[(\frac{p_g}{m_i} + e_g)^{\frac{\rho-1}{\rho}} - (\frac{p_b}{m_i} + e_b)^{\frac{\rho-1}{\rho}}\right]$. Note that $\rho > 1$ implies $\frac{\rho}{1-\rho} < 0$ and $\frac{\rho-1}{\rho} \in (0, 1)$. Thus, as $m_i \to \infty$, (a) $\frac{\rho}{1-\rho} \to \frac{\rho}{1-\rho} < 0$, (b) $m_i^{\frac{\rho-1}{\rho}} \to \infty$, and (c) $\left[(\frac{p_g}{m_i} + e_g)^{\frac{\rho-1}{\rho}} - (\frac{p_b}{m_i} + e_b)^{\frac{\rho-1}{\rho}}\right] \to e_g^{\frac{\rho-1}{\rho}} - e_b^{\frac{\rho-1}{\rho}} < 0$. Consequently, Condition 4 holds. □

Proof of Proposition 2.2:

Proof. Let M^b be the set of all m_i s.t. i weakly prefers to adopt the brown technology. Formally, $M^b = \{m_i \mid \Delta^{use}(m_i) \leq \Delta^{fix}\}$. By Assumption 2.1, $\min[\Delta^{use}(m_i)] \leq \Delta^{fix} \Rightarrow \exists m_i$ s.t. $\Delta^{use}(m_i) \leq \Delta^{fix} \Rightarrow M^b \neq \{\emptyset\}$. Next, suppose that M^b was not a connected set. Hence, there is $x < y < z$ s.t. $x, z \in M^b$ but $y \notin M^b$. Hence, $\Delta^{use}(x) \leq \Delta^{fix}$, $\Delta^{use}(y) \geq \Delta^{fix}$, and $\Delta^{use}(z) \leq \Delta^{fix}$. Thus, $\Delta^{use}(x) < \Delta^{use}(y)$ and $\Delta^{use}(z) < \Delta^{use}(y)$. But this implies that there is a local maximum $a \in (x, z)$, which contradicts Lemma 2.1, which implies that $\Delta^{use}(m_i)$ has at most one extremum, which must be a minimum. □

Proof of Lemma 2.2:

Proof. Assumption 2.1 implies that there exists at least one m such that $\Delta^{fix} = \Delta^{use}(m)$. Moreover, since $\Delta^{fix} > \Delta^{use}(0)$ and $\Delta^{use}(m)$ has at most one extremum, which must be a minimum, this intersection is unique. We denote it by τ. Because $\lim_{m\to\infty} \Delta^{use}(m) = \infty$, it must hold that $\left.\frac{\partial \Delta^{use}(m)}{\partial m}\right|_{m=\tau} > 0$. By Lemma 2.1, this implies $\tau > \bar{m}$. Moreover, form the above follows that $\Delta^{fix} > \Delta^{use}(x)$ if and only if $x > \tau$, which by definition of $M^b(\Delta^{fix})$ implies $\tau = \max M^b(\Delta^{fix})$. Since $\Delta^{use}(m)$ is strictly increasing for all $m > \bar{m}$, it follows that $\Delta^{fix} < \Delta^{use}(m)$ for all $m > \tau$. Since τ is unique, $\Delta^{fix} > \Delta^{use}(m)$ for all $m < \tau$, which implies $\min M^b(\Delta^{fix}) = 0$. By continuity of $\Delta^{use}(m)$, there exists $\delta > 0$ such that for all $\epsilon \in (-\delta, \delta)$, $\Delta^{fix} + \epsilon > \Delta^{use}(0)$. For any such ϵ, the above reasoning applies verbatim, implying $\min M^b(\Delta^{fix} + \epsilon) = 0$ and therefore $\frac{\partial \min M^b(\Delta^{fix})}{\partial \Delta^{fix}} = 0$. Total emissions can be differentiated as $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} = \frac{\partial \max M^b(\Delta^{fix})}{\partial \Delta^{fix}} f(\tau) \left[x^*(\tau, b)e_b - x^*(\tau, g)e_g\right]$. Since $\tau > \bar{m}$, we have $x^*(\tau, b)e_b - x^*(\tau, g)e_g > 0$, and because $f(\tau) \geq 0$, it follows that $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \geq 0 \iff \frac{\partial \max M^b(\Delta^{fix})}{\partial \Delta^{fix}} \geq 0$. To determine the sign of this derivative, consider the defining equality $\frac{\rho}{1-\rho}\left[(p_g + \max M^b(\Delta^{fix})\, e_g)^{\frac{\rho-1}{\rho}} - (p_b + \max M^b(\Delta^{fix})\, e_b)^{\frac{\rho-1}{\rho}}\right] = \Delta^{fix}$. Differentiating both sides with respect to Δ^{fix} and rearranging yields $\frac{\partial \max M^b(\Delta^{fix})}{\partial \Delta^{fix}} = \frac{1}{e_b\,(p_b + \max M^b(\Delta^{fix})\, e_b)^{-1/\rho} - e_g\,(p_g + \max M^b(\Delta^{fix})\, e_g)^{-1/\rho}} = \frac{1}{e_b\, x^*(\tau, b) - e_g\, x^*(\tau, g)}$. Because $\tau > \bar{m}$, the denominator is strictly positive, implying $\frac{\partial \max M^b(\Delta^{fix})}{\partial \Delta^{fix}} > 0$ and thus $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \geq 0$. □

Proof of Lemma 2.3:

Proof. Assumption 2.1 implies that there exists at least one m such that $\Delta^{fix} = \Delta^{use}(m)$. Moreover, since (1) $\Delta^{fix} < \Delta^{use}(0)$, (2) $\Delta^{use}(m)$ has at most one extremum, which must be a minimum, and (3) $\lim_{m\to\infty} \Delta^{use}(m) = \infty > \Delta^{fix}$, there are two intersections. We denote them by $\tau_l < \tau_h$. Because $\lim_{m\to\infty} \Delta^{use}(m) = \infty$, it must hold that $\left.\frac{\partial \Delta^{use}(m)}{\partial m}\right|_{m=\tau_h} > 0$ and $\left.\frac{\partial \Delta^{use}(m)}{\partial m}\right|_{m=\tau_l} < 0$. By Lemma 2.1, this implies $\tau_l < \bar{m} < \tau_h$. Since $\Delta^{use}(m)$ has at most one extremum, which must be a minimum, it must hold that $x \in M^b(\Delta^{fix}) \Leftrightarrow x \in [\tau_l, \tau_h]$. By definition of $M^b(\Delta^{fix})$, we have $\tau_l = \min[M^b(\Delta^{fix})]$ and $\tau_h = \max[M^b(\Delta^{fix})]$. Total emissions can be differentiated as $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} = \frac{\partial \min[M^b(\Delta^{fix})]}{\partial \Delta^{fix}} f(\tau_l) \left[x^*(\tau_l, g)e_g - x^*(\tau_l, b)e_b\right] + \frac{\partial \max[M^b(\Delta^{fix})]}{\partial \Delta^{fix}} f(\tau_h) \left[x^*(\tau_h, b)e_b - x^*(\tau_h, g)e_g\right]$.

By analogous reasoning as in the proof of Lemma 2.2 (considering the defining equality of τ_l and τ_h, and taking derivatives with respect to Δ^{fix}), we obtain $\frac{\partial \max[M^b(\Delta^{fix})]}{\partial \Delta^{fix}} = \frac{1}{e_b\, x^*(\tau_h, b) - e_g\, x^*(\tau_h, g)}$ and $\frac{\partial \min[M^b(\Delta^{fix})]}{\partial \Delta^{fix}} = \frac{1}{e_b\, x^*(\tau_l, b) - e_g\, x^*(\tau_l, g)}$. Substituting into the derivative of $E(\Delta^{fix})$ yields $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} = f(\tau_h) - f(\tau_l)$, from which the lemma follows directly. □

Proof of Lemma 2.4:

Proof. Suppose $\frac{\partial E(\Delta^{fix})}{\partial \Delta^{fix}} \leq 0$. The proof of Lemma 2.3 shows that there is τ_l, τ_h s.t. $\tau_l < \bar{m} < \tau_h$ and $M^b = [\tau_l, \tau_h]$. Consider any $y \in (\bar{m}, \tau_h)$. Hence, (1) $y \in M^b(\Delta^{fix})$ and (2) $y > \bar{m} \Rightarrow e_g x^*(y, g) < e_b x^*(y, b)$, which proves the lemma. □