Eco-Certification: Warm Glow or Cold Prickle?

Klaas van 't Veld
Department of Economics, University of Wyoming

Abstract

Eco-certification programs, which allow firms to display an eco-label on a product if it meets some environmental quality standard, have become a prominent complement to, and sometimes substitute for, environmental policies. This paper analyzes two common approaches to modeling why consumers are willing to pay extra for certified products: one approach has consumers experience positive utility from benefiting the environment; the other has them experience negative utility from damaging the environment. The two approaches are shown to be fully equivalent if consumers' utility weights and reference points for either benefits or damages are exogenous. Plausibly, however, these weights and reference points are endogenous, both to policies such as environmental awareness campaigns and to the behavior of other consumers. This endogeneity can lead to radically different welfare implications and policy recommendations.

1. Introduction

Eco-certification programs, which allow firms to display an eco-label on a product if it meets some environmental quality standard, have become a prominent complement to, and sometimes substitute for, environmental policies. The theoretical literature analyzing these programs has thus far focused mainly on how economic and environmental outcomes in so-called green markets are shaped by interactions between certifiers, regulators, and firms. Far less attention has been paid to the consumer side, which has typically been modeled very simply, using one of two approaches.

Both approaches typically have consumers compare a "green," eco-certified good to a "brown," non-certified good, whereby the green good is more expensive. The difference lies in how consumer utility is modeled. Under the "benefit" approach, each consumer's utility is treated as increasing in the environmental benefit that buying the green good provides, relative to the lower or zero environmental benefit from buying the brown good. Under the "avoided-damage" approach, utility is instead treated as decreasing in the environmental damage that buying the brown good causes, relative to the lower or zero damage from buying the green good.

It is not hard to show that both modeling approaches are fully equivalent, in terms of not just market and environmental outcomes but also welfare effects, provided four conditions are met.

First, units must obviously be aligned: one unit of environmental benefits must correspond to one unit of avoided damages. Second, benefits and damages must be measured relative to a consistent and fixed baseline. Third, and distinct from the second condition, consumers must *frame* benefits and damages relative to a consistent and fixed baseline. And, fourth, consumers' utility weights on either benefits or damages must be treated as fixed.

This paper is concerned with the third and fourth conditions, how they interact, and what happens if they fail to hold. As pointed out in a recent review of the eco-labeling literature (van 't

Veld, 2020), plenty of empirical evidence suggests that consumers do *not* have fixed frames and utility weights when making choices in green markets.[1]

Coming to grips with this evidence requires asking the fundamental question of why exactly consumers are willing to pay sometimes hefty premia for eco-certified goods. The literature has mostly punted on this question, simply taking the higher willingness to pay as given. Occasionally, reference is made to the concept of "warm glow," which was introduced by Andreoni (1989) to explain empirical evidence on charitable giving. Just like charitable donors may get a positive warm-glow feeling from the very act of "doing their bit," green consumers may get warm glow from the very act of purchasing an environmentally friendly good.

Andreoni himself has noted, however, that the concept of warm glow is really just a placeholder for more specific models of why people donate to charity. One such model is that people give not to feel good, but rather to avoid feeling bad, i.e., avoid feeling guilty about *not* giving. Andreoni (1995) refers to such guilt as "cold prickle" and shows that, depending on how a public-goods game is framed in the lab, either warm glow from contributing to others or avoiding cold prickle from free-riding can be activated in participants' minds, with very different outcomes.

A recent informal analysis by Miller and Monin (2016) expands on this, noting that many behavioral phenomena involving moral choices can be unified by assuming that people divide such choices into two categories. In situations perceived as moral *opportunities*, choosing the moral option enhances one's moral self-image, giving rise to warm glow, whereas not choosing the option leaves one's self-image unchanged. In contrast, in situations perceived as moral *tests*, failing to choose the moral option diminishes one's moral self-image, giving rise to cold prickle, whereas choosing the option leaves one's self-image unchanged. Miller and Monin emphasize, moreover, that these perceptions are context-dependent: "What constitutes a test for one person in one context may not for another person or for the same person in another context. The same is true of opportunities. (p. 45)"

In a green-market setting, this suggests that, whereas some consumers may view an eco-labeled product as a moral opportunity—a potential source of warm glow if they buy it—other consumers may view the product as a moral test—a potential source of guilt if they fail to buy it. In essence, the former consumers focus on environmental benefits, while the latter focus on environmental damages. Even so, as long as the four conditions listed above hold—in particular, as long as consumers apply fixed utility weights and reference points to either benefits or damages—their differing frames turn out to have no implications for market, environmental, or even welfare outcomes.

What happens, however, if consumers' utility weights change? Consider, as a simple example, an environmental awareness campaign aimed at increasing consumers' concern for the environment, as captured by their utility weights. I find that under reasonable assumptions such a campaign induces behavior switches that unambiguously improve environmental outcomes, even though (by shifting some consumers to abstaining from consuming the good altogether) the campaign may paradoxically reduce purchases of eco-labeled goods. Moreover, these behavioral and environmental effects are identical regardless of what moral frame consumers adopt, i.e., regardless of whether they focus on environmental benefits or damages.

[1] The ideas developed formally in this paper were sketched out informally or hinted at in that review.

The same is not true of the campaign's welfare effects, however. If most consumers adopt a moral-opportunity frame—treating consumption of a brown good as their utility-reference point, and experiencing warm glow from any behavior greener than that—then raising their utility weights raises their warm glow. Since the resulting improvement in environmental outcomes also enhances public benefits, the campaign unambiguously increases welfare. In contrast, if most consumers adopt a moral-test frame—treating green consumption or even abstention from consumption as their utility-reference point, and experiencing guilt from any behavior browner than that—then raising their utility weights raises their guilt. Any public environmental gains from the campaign must then be weighed against consumers' private utility losses, and the net welfare effect may well be negative.

More generally, once moral framing is taken into account, and particularly when the utility reference point relative to which consumers experience either warm glow or cold prickle is endogenized in plausible ways, ambiguity of welfare effects becomes pervasive.

Consider, for example, the introduction of a green good into a market where initially all consumers buy a brown good. Absent moral-framing effects, this should always at least weakly increase private utility, since consumers always have the option of sticking with their current choice. I find, however, that once moral-framing effects are taken into account, the initial private welfare increase may well fade over time, as norms of behavior adjust to the new market reality. In particular, if the green good captures a large market share, buying green may become the "new normal," relative to which consumers start measuring their behavior. If so, this may spark new feelings of guilt in consumers who stuck with buying brown, while also eroding the warm glow felt by consumers who switched to green. In fact, I find for a benchmark case with uniformly distributed environmental awareness that these changing norms may turn the short-run welfare gain into a long-run welfare *loss*.

This is only one of several possible outcomes, however. In fact, the exact opposite may happen if a green good is introduced in a setting where no good existed before. Since consuming the new good will likely still harm the environment to some extent, highly eco-conscious consumers may choose to abstain from buying it, while less eco-conscious consumers, who do choose to buy it, may end up feeling a certain amount of guilt. However, if buying the green good eventually becomes the new normal, this may spark new feelings of warm glow in abstainers, while also eroding the guilt felt by green consumers. These long-run "norm effects" will then reinforce the short-run private utility gains from introducing the green good, helping offset any negative environmental impacts.

An intermediate case arises when a green good is introduced in a market where initially some consumers buy a brown good, while others abstain from consumption, to avoid the brown good's damages. In this setting, *both* types of norm effects described above come into play. Surprisingly, in the benchmark case, they turn out to offset each other exactly, thereby leaving the short-run private welfare increase unchanged.

All the above results consider a green good of arbitrary environmental quality and price, varying only consumer awareness or the green good's presence in a market. In practice, of course, environmental quality and price are related, since producing higher quality is costly. Moreover, because environmental quality is usually a "credence" attribute—one that consumers cannot verify individually—firms can hope to recover their higher costs only if they can find an eco-certification program willing to vouch for their product's quality. As a result, the actual quality provided by firms is effectively determined by the performance standards that certification programs choose to certify to.

A common question examined in the theoretical literature on eco-certification is how these standards are determined, and how they might vary depending on certification programs' objectives. Several studies show that programs sponsored by profit-maximizing industry associations set the lowest standards, programs sponsored by environmental-benefit maximizing NGOs set the highest standards, and programs sponsored by welfare-maximizing governments set standards in between. None of these studies allow for norm effects, however.

This paper's finding that norm effects might reduce welfare from introducing a green good, particularly if the good captures a large market share, suggests that a welfare-maximizing government sponsor might purposely shade its standard to limit the good's popularity, and thereby mitigate negative norm effects. Indeed, this turns out to be the case. When I extend the model of Li and van 't Veld (2015) to incorporate norm effects, I find that a government sponsor may not just optimally set a standard weaker than that of an industry sponsor, but may even optimally abstain from sponsoring a green good in the first place.

The remainder of the paper is structured as follows. Section 2 introduces two models of a green market, one using the benefit approach and one using the avoided-damage approach, and shows these to be fully equivalent under the four conditions spelled out above. Section 3 then breaks that equivalence, by considering the welfare effects of an awareness campaign that raises consumers' utility weights. Section 4 breaks the equivalence further, by considering plausible changes over time in consumer norms relative to which warm glow and guilt are measured. Section 5 considers the implications for optimal standard-setting by a government, and Section 6 concludes.

2. Model

Overwhelmingly, the theoretical literature on eco-label programs has modeled consumer preferences using the framework introduced by Mussa and Rosen's (1978) analysis of optimal pricing and quality provision by monopolist firms. In that framework, each consumer buys a single unit of a product and receives indirect utility $\theta q - p$, where q is the product's quality, p its price, and θ a parameter measuring the intensity of the consumer's taste for quality.

In Mussa and Rosen's original model, the precise definition of "quality" is arbitrary, as is its measurement; the authors note that they could make utility concave in q, but that would merely involve a redefinition of the units in which quality is measured. In applications of their model to eco-label settings, however, q is typically treated as some objective measure of the environmental "benefit" associated with a product. This benefit is moreover treated as a credence quality, meaning that it is prohibitively costly for individual consumers to verify how large q is. This is where eco-labeling organizations come in: they do the work of certifying q, and then allowing firms to credibly signal q by affixing an eco-label to their products.

Consider, then, a typical model of a market with N potential consumers in which two versions of a good coexist: a brown version that does not generate any environmental benefit, and a green version that does, but that as a result is more expensive. Let

$$U^b = u - p^b$$

denote a consumer's private utility from consuming a unit of the brown good, where u is the good's use value—derived from its taste, convenience, etc.—and p^b is its price. Also, let

$$U^g = u + \theta q - p^g$$

denote the utility from consuming the green good, which has the same use value but additionally provides environmental benefit $q \in (0,1)$ weighted by θ, and which costs $p^g > p^b$. As this section will show, this environmental benefit can be conceptualized equally well as an avoided environmental damage d, where $d = 1$ represents the (normalized) maximal damages caused by consuming the brown version, and $d = 1 - q$ the lower, but still positive, damages caused by consuming the green version. Lastly, let

$$U^a = \theta.$$

denote the consumer's utility if she abstains from consuming either version. By doing so, she forgoes the use value u, but saves on the price p^g or p^b, and also avoids all damages associated with consumption, so $d = 0$. Equivalently, her abstaining provides the maximal environmental benefit, $q = 1$.

The full utility that consumers get from each of these three choices also includes a public component βQ, equal to the aggregate environmental benefits $Q = N^a \cdot 1 + N^g \cdot q$ provided by the N^a abstainers and N^g green consumers combined, weighted by parameter β. Because these benefits are public, the N^b brown consumers get to enjoy them as well, despite not contributing to them. Because this component is therefore constant across the three choices, it will play a role only in later sections, where we consider aggregate welfare analysis. In this section, we will suppress it, taking it as understood.

To reduce notational clutter, let $p \equiv p^g - p^b$ denote the price *premium* fetched by the green good, and $v \equiv u - p^b$ as the use value of the good net of the brown good's price. With these normalizations, the above equations simplify to

$$U^b = v$$
$$U^g = v + \theta q - p$$
$$U^a = \theta.$$

To introduce consumer heterogeneity, treat weight θ as a declining function of an index variable n on $[0, N]$. That then makes it easy to think of ordering consumers by declining θ, by lining them up in order of that index variable. The "first" consumer, with index $n = 0$, will have the highest utility weight $\overline{\theta} \equiv \theta(0)$, while the "last" consumer, with index $n = N$, will have utility weight $\theta(N)$ normalized to zero.

Several of the results below depend on the shape of the weight function $\theta(n)$. This shape in turn depends on how the utility weight θ is distributed in the population. If the distribution of θ

is uniform—a simple benchmark case—then the $\theta(n)$ function is linear, and can be written as

$$\theta(n) = \overline{\theta}\left(1 - \frac{n}{N}\right).$$

By way of setup for the analysis to follow, it is useful to introduce some additional notation, separately for three benchmark cases: (1) a market in which initially all consumers buy the brown product, (2) a degenerate "market" in which initially all consumers abstain from buying the product, and (3) a mixed market in which initially all consumers either abstain or buy brown. Throughout, we will break in difference in favor of abstaining over buying, and buying green over buying brown.

2.1. **Initial all-brown market**

Suppose then that initially no green good exists, and that all N consumers buy brown rather than abstaining. For this to be optimal requires that $\overline{\theta} < v$, since then $U^a < U^b \Leftrightarrow \theta(n) < v$ for all consumers, including the very greenest.

If now a green good is introduced, some consumers will buy it, provided its environmental benefit q and price premium p are such that $\overline{\theta}q \geq p$: the private utility benefit $\overline{\theta}q$ gained by the greenest consumer from buying it must outweigh the higher cost p. If this condition holds, all consumers for whom

$$\begin{aligned} & U^g \geq U^b \\ \Leftrightarrow \quad & v + \theta(n)q - p \geq v \\ \Leftrightarrow \quad & \theta(n) \geq \frac{p}{q} \equiv \theta^{gb}. \end{aligned}$$

will buy green. It will be useful to let n^{gb} denote the index of the last consumer to do so, implicitly defined by

$$\theta(n^{gb}) = \frac{p}{q},$$

and to let $\mathcal{N}^g_{gb}$ denote the set of indices $n \in [0, n^{gb}]$ of the N^g_{gb} consumers that buy the green good. Similarly, let $\mathcal{N}^b_{gb}$ denote the set of indices $n \in (n^{gb}, N]$ of the remaining $N^b_{gb} = N - N^g_{gb}$ consumers that stick with buying brown. Some consumers are guaranteed to do so, since by assumption the least-green consumer has utility weight $\theta(N) = 0$, which is less than p/q.

As a result of the green good's introduction, aggregate environmental benefits change from $Q^b = 0$ to $Q^{gb} = N^g_{gb} \cdot q$, so environmental quality improves by

$$Q^{gb} - Q^b = N^g_{gb} \cdot q > 0. \tag{1}$$

The public component of welfare therefore increases by $B(Q^{gb} - Q^b) = BN^g_{gb} \cdot q$, where $B \equiv N\beta$ is the aggregate utility weight on the public utility benefits for all consumers combined.

Meanwhile, private welfare—really just consumer surplus, since we ignore any producer surplus for now—changes from

$$W^b = \int_{\mathcal{N}} \underbrace{v}_{U^b}\, dn$$

to

$$W^{gb} = \int_{\mathcal{N}_{gb}^g} \underbrace{[v + \theta(n)q - p]}_{U^g} \, dn + \int_{\mathcal{N}_{gb}^b} \underbrace{v}_{U^b} \, dn,$$

so improves by

$$W^{gb} - W^b = \int_{\mathcal{N}_{gb}^g} \underbrace{[\theta(n)q - p]}_{U^g - U^b > 0} \, dn > 0.$$

2.2. **Initial all-abstention market**

Suppose next that initially neither a brown nor a green good exist, so that trivially all consumers "abstain" from consumption. The very first good introduced, thereby *creating* a market, just happens to be green.

Some consumers will end up buying that green good as long as $v > p$, because then for the least green consumer, with utility weight $\theta(N) = 0$, we have that

$$\begin{aligned} & U^g > U^a \\ \Leftrightarrow \quad & v + \theta(n)q - p > \theta(n). \end{aligned}$$

Conversely, as long as $\overline{\theta}(1 - q) \geq v - p$, at least some consumers will continue to abstain, namely all consumers with utility weights such that

$$\begin{aligned} & U^a \geq U^g \\ \Leftrightarrow \quad & \theta(n) \geq v + \theta(n)q - p \\ \Leftrightarrow \quad & \theta(n) \geq \frac{v - p}{1 - q} \equiv \theta^{ag}. \end{aligned}$$

It will be useful to let n^{ag} denote the index of the last consumer to abstain, implicitly defined by

$$\theta(n^{ag}) = \frac{v - p}{1 - q},$$

and to let $\mathcal{N}_{ag}^a$ denote the set of indices $n \in [0, n^{ag}]$ of the N_{ag}^a consumers that stick with abstention. Similarly, let $\mathcal{N}_{ag}^g$ denote the set of indices $n \in (n^{ag}, N]$ of the remaining $N_{ag}^g = N - N_{ag}^a$ consumers that switch to buying green.

Aggregate environmental benefits change from

$$Q^a = N = N_{ag}^a \cdot 1 + N_{ag}^g \cdot 1$$

to

$$Q^{ag} = N_{ag}^a \cdot 1 + N_{ag}^g \cdot q,$$

so in this case they *fall* by

$$Q^{ag} - Q^a = -N_{ag}^g \cdot (1 - q) < 0. \tag{2}$$

Meanwhile, welfare changes from

$$W^a = \int_{\mathcal{N}} \underbrace{\theta(n)}_{U^a} \, dn$$

to

$$W^{ag} = \int\limits_{\mathcal{N}_{ag}^{a}} \underbrace{\theta(n)}_{U^a} \, dn + \int\limits_{\mathcal{N}_{ag}^{g}} \underbrace{[v + \theta(n)q - p]}_{U^g} \, dn,$$

so improves by

$$W^{ag} - W^{a} = \int\limits_{\mathcal{N}_{ag}^{g}} \underbrace{[v - \theta(n)(1-q) - p]}_{U^g - U^a > 0} \, dn > 0.$$

2.3. **Initial mixed market**

Suppose, as a third possibility, that the market initially, before introducing a green good, has a mix of abstainers and brown-good buyers. This requires that $\overline{\theta} \geq v$, since then $U^a \geq U^b$ for all consumers with utility weights

$$\theta(n) \geq v = \theta^{ab}. \tag{3}$$

Let n^{ab} denote the index of the last consumer to abstain, implicitly defined by

$$\theta(n_{ab}^{a}) = v.$$

Also, let $\mathcal{N}_{ab}^{a}$ denote the set of indices $n \in [0, n^{ab}]$ of the N_{ab}^{a} initial abstainers, and $\mathcal{N}_{ab}^{b}$ the set of indices $n \in (n^{ab}, N]$ of the remaining $N_{ab}^{b} = N - N_{ab}^{a}$ consumers that initially buy brown. Some such consumers are guaranteed to exist, since $\theta(N) = 0 < v$.

If now a green good is introduced, consumers for whom both $U^g \geq U^b$ and $U^g > U^a$ will buy it. From the analysis of the previous sub-sections, this will be true for consumers with utility weights

$$\theta(n) \in \left[\frac{p}{q}, \frac{v-p}{1-q}\right) \equiv \left[\theta^{gb}, \theta^{ag}\right).$$

For this interval to be non-empty requires that

$$\begin{aligned} \frac{v-p}{1-q} &> \frac{p}{q} \\ \Leftrightarrow \qquad vq &> p. \end{aligned}$$

To understand this condition, note from (3) above that the marginal consumer who is just indifferent between abstaining or buying brown before the green good is introduced has utility weight $\theta^{ab} = v$. The condition can therefore also be written as $\theta^{ab} q > p$, showing that, for the green good to gain any traction in the market, the utility benefit gained by this marginal consumer from switching to green must strictly outweigh the green good's price premium.

If this condition holds, then of the initial N_{ab}^{a} consumers who abstained, N_{agb}^{a} consumers, with indices $n \in [0, n^{ag}]$, will continue to do so. Similarly, of the initial N_{ab}^{b} consumers who bought brown, N_{agb}^{b} consumers, with indices $n \in (n^{gb}, N]$ will continue to do so. Let $\mathcal{N}_{agb}^{a}$ and $\mathcal{N}_{agb}^{b}$ denote the corresponding index sets, and $\mathcal{N}_{agb}^{g}$ the set of indices in $(n^{ag}, n^{gb}]$ of the N_{agb}^{g} consumers who switch to buying green.

Of these new green consumers, $N_{agb}^{a\to g} \equiv N_{ab}^{a} - N_{agb}^{a}$ will have switched from previously abstaining, while $N_{agb}^{b\to g} \equiv N_{ab}^{b} - N_{agb}^{b}$ will have switched from previously consuming brown. Let $\mathcal{N}_{agb}^{a\to g}$ and $\mathcal{N}_{agb}^{b\to g}$ denote the corresponding index sets $(n^{ag}, n^{ab}]$ and $(n^{ab}, n^{gb}]$.

The green good's introduction changes aggregate environmental benefits from

$$Q^{ab} = N_{ab}^{a} \cdot 1 + N_{ab}^{b} \cdot 0$$

to

$$\begin{aligned} Q^{agb} &= N_{agb}^{a} \cdot 1 + N_{agb}^{g} \cdot q + N_{agb}^{b} \cdot 0 \\ &= N_{agb}^{a} \cdot 1 + N_{agb}^{a\to g} \cdot q + N_{agb}^{b\to g} \cdot q + N_{agb}^{b} \cdot 0, \end{aligned}$$

so environmental quality changes by

$$Q^{agb} - Q^{ab} = -N_{agb}^{a\to g} \cdot (1-q) + N_{agb}^{b\to g} \cdot q \gtreqless 0. \tag{4}$$

In general, this change is ambiguous in sign. On the one hand, the $N_{agb}^{a\to g}$ switchers to green from abstaining start causing damages $d = 1 - q$, which reduces environmental quality. On the other hand, the $N_{agb}^{b\to g}$ switchers to green from brown reduce their damages from $d = 1$ to $d = 1 - q$, which improves environmental quality. In general, either effect might dominate.

Meanwhile, welfare changes from

$$W^{ab} = \int_{\mathcal{N}_{ab}^{a}} \underbrace{\theta(n)}_{U^a}\, dn + \int_{\mathcal{N}_{ab}^{b}} \underbrace{v}_{U^b}\, dn$$

to

$$\begin{aligned} W^{agb} &= \int_{\mathcal{N}_{agb}^{a}} \underbrace{\theta(n)}_{U^a}\, dn + \int_{\mathcal{N}_{agb}^{g}} \underbrace{[v + \theta(n)q - p]}_{U^g}\, dn + \int_{\mathcal{N}_{agb}^{b}} \underbrace{v}_{U^b}\, dn \\ &= \int_{\mathcal{N}_{agb}^{a}} \underbrace{\theta(n)}_{U^a}\, dn + \int_{\mathcal{N}_{agb}^{a\to g}} \underbrace{[v + \theta(n)q - p]}_{U^g}\, dn + \int_{\mathcal{N}_{agb}^{b\to g}} \underbrace{[v + \theta(n)q - p]}_{U^g}\, dn + \int_{\mathcal{N}_{agb}^{b}} \underbrace{v}_{U^b}\, dn, \end{aligned}$$

so unambiguously improves by

$$W^{agb} - W^{ab} = \int_{\mathcal{N}_{agb}^{a\to g}} \underbrace{[v - \theta(n)(1-q) - p]}_{U^g - U^a > 0}\, dn + \int_{\mathcal{N}_{agb}^{b\to g}} \underbrace{[\theta(n)q - p]}_{U^g - U^b > 0}\, dn > 0.$$

2.4. **Avoided-damage approach**

With the setup and notation established in the previous three subsections, it is straightforward to show that, as long as units and baselines are treated consistently and utility weights are held constant throughout, adopting an avoided-damage approach, in which consumers experience negative private utility from environmental damages caused rather than positive privately utility from environmental benefits generated, is fully equivalent in terms of market, environmental, and welfare outcomes.

Under the damage approach a consumer who buys brown enjoys the good's use value v (net of the brown good's price p^b), but gets disutility $\theta \cdot 1$ from causing maximal environmental damages $d = 1$:

$$\widetilde{U}^b = v - \theta.$$

A consumer who buys green enjoys the same use value, but gets lower disutility $\theta \cdot d$, since the green good's damages $d = 1 - q$ are lower. In addition, she incurs the green good's price premium p, so

$$\widetilde{U}^g = v - \theta d - p.$$

Lastly, a consumer who abstains from consuming either version of the good forgoes the use value, saves on the price, and also avoids all disutility, because she causes no damages. Her utility is therefore

$$\widetilde{U}^a = 0.$$

The full utility also includes a public component βD, equal to the aggregate environmental damages $D = N^g \cdot d + N^b \cdot 1$ caused by the N^g green and N^b brown consumers combined, weighted by parameter β.

In an initial all-brown market, consumers with these utility functions will switch to buying a newly introduced green good if

$$\begin{aligned} & \widetilde{U}^g \geq \widetilde{U}^b \\ \Leftrightarrow \quad & v - \theta(n)d - p \geq v - \theta(n) \\ \Leftrightarrow \quad & \theta(n) \geq \frac{p}{1-d} = \frac{p}{q} \equiv \theta^{gb}. \end{aligned}$$

This condition is identical to that under the benefit approach, and therefore applies to the same N^g_{gb} consumers.

Environmental damages change from

$$D^b = N \cdot 1$$

to

$$D^{gb} = N^g_{gb} \cdot d + N^b_{gb} \cdot 1,$$

so environmental quality improves by

$$D^b - D^{gb} = N^g_{gb} \cdot (1 - d) = N^g_{gb} \cdot q = Q^b - Q^{gb}.$$

This, too, is exactly the same change as under the benefit approach.

Lastly, welfare changes from

$$\widetilde{W}^b = \int_{\mathcal{N}} [v - \theta(n)]\, dn$$

to

$$\widetilde{W}^{gb} = \int_{\mathcal{N}^g_{gb}} [v - \theta(n)d - p]\; dn + \int_{\mathcal{N}^b_{gb}} [v - \theta(n)]\, dn,$$

so improves by

$$\widetilde{W}^{gb} - \widetilde{W}^{b} = \int\limits_{\mathcal{N}^g_{gb}} \underbrace{[\theta(n)(1-d) - p]}_{\widetilde{U}^g - \widetilde{U}^b > 0}\, dn = \int\limits_{\mathcal{N}^g_{gb}} \underbrace{[\theta(n)q - p]}_{\widetilde{U}^g - \widetilde{U}^b > 0}\, dn = W^{gb} - W^{b}.$$

Once again, this identical to the change under the benefit approach.

It is straightforward to check that same equivalence holds for the other two market scenarios examined in subsections 2.2 and 2.3. Key to the equivalence is that, when we compare utilities under the two approaches,

$$\begin{aligned} U^b &= v & \qquad & \widetilde{U}^b = v - \theta \\ U^g &= v + \theta q - p & \text{vs.} \qquad & \widetilde{U}^g = v - \theta d - p \\ U^a &= \theta & \qquad & \widetilde{U}^a = 0, \end{aligned}$$

we see that the avoided-damage approach simply subtracts θ from all three benefit-approach utilities. That then leaves the critical weights θ^{ag}, θ^{ab}, and θ^{gb} at which consumers are indifferent between various choices unchanged, and thereby also the critical indices n^{ag}, n^{ab}, and n^{gb}, as well as the implied sets of abstainers, green consumers, and brown consumers in the different scenarios. Moreover, because private welfare under the avoided-damage approach equals that under the benefit approach up to a constant term

$$-\int\limits_{\mathcal{N}} \theta(n)\, dn,$$

private-welfare comparisons are identical as well.

As for comparisons of environmental quality, since aggregate environmental benefits can be written as

$$\begin{aligned} Q &= N^a \cdot 1 + N^g \cdot q + N^b \cdot 0 \\ &= N - [N^a \cdot 0 + N^g \cdot (1-q) + N^b \cdot 1] \\ &= N - D, \end{aligned}$$

increases in Q translate into equivalent decreases in D. That then also implies that the public components of utility, $+\beta Q$ under the benefit approach and $-\beta D$ under the avoided-damage approach, are equivalent, up to constant term N.[2]

3. Awareness campaigns

The straightforward equivalence of the benefit and avoided-damage approaches suggests that either approach is equally valid when analyzing green markets. The choice between them comes down to nothing more than a different normalization, using a different baseline.

It turns out, however, that the equivalence breaks down, and does so quite thoroughly, when analyzing policies that have the effect of altering consumers' utility weights on benefits or damages.

[2] Plausibly, the public components are concave in Q and convex in D, written for example as $+\beta g(Q)$, with $g' > 0$, $g'' < 0$, and $-\beta h(D)$, with $h' > 0$, $h'' > 0$. Equivalence then requires that $g(Q) = -h(N-Q)$ and therefore $g(N-D) = -h(D)$.

The most straightforward example of such a policy is an awareness campaign, aimed at increasing consumers' concern for the environment.

A simple way to model the effects of such a campaign is to take the benchmark case where utility weights $\theta(n)$ are linear, of the form $\theta(n) = \overline{\theta}(1 - n/N)$, and suppose that the campaign has the effect of increasing parameter $\overline{\theta}$.[3]

In this benchmark case, we can solve explicitly for the index n^{gb} of a consumer who is just indifferent between buying green or brown:

$$\overline{\theta}\left(1 - \frac{n^{gb}}{N}\right) = \theta^{gb}$$
$$\Leftrightarrow \quad n^{gb} = N\left(1 - \frac{\theta^{gb}}{\overline{\theta}}\right).$$

In the case of a so-called "covered" market, with no abstainers, the number of green consumers equals that index, i.e., $N^g_{gb} = n^{gb}$. We therefore have

$$\frac{\partial N^g_{gb}}{\partial \overline{\theta}} = \frac{\partial n^{gb}}{\partial \overline{\theta}} = \frac{N}{\overline{\theta}^2} \cdot \theta^{gb} = \frac{N}{\overline{\theta}^2} \cdot \frac{p}{q} > 0.$$

That is, an environmental awareness campaign will increase the number of green consumers in a covered market. It will thereby also improve aggregate environmental quality,

$$Q^{gb} = N^g_{gb} \cdot q + [N - N^g_{gb}] \cdot 0,$$

namely by

$$\frac{\partial Q^{gb}}{\partial \overline{\theta}} = \frac{\partial N^g_{gb}}{\partial \overline{\theta}} \cdot q = \frac{N}{\overline{\theta}^2} \cdot p > 0.$$

Similarly, in a market with only abstainers and green consumers, the index n^{ag} of the marginal consumer is

$$n^{ag} = N\left(1 - \frac{\theta^{ag}}{\overline{\theta}}\right),$$

and since the number of abstainers equals that index, $N^a_{ag} = n^{ag}$, we have

$$\frac{dN^a_{ag}}{d\overline{\theta}} = \frac{dn^{ag}}{d\overline{\theta}} = \frac{N}{\overline{\theta}^2} \cdot \theta^{ag} = \frac{N}{\overline{\theta}^2} \cdot \frac{v - p}{1 - q} > 0.$$

An environmental awareness campaign will therefore increase the number of abstainers, and thereby again improve aggregate environmental quality

$$Q^{ag} = N^a_{ag} \cdot 1 + [N - N^a_{ag}] \cdot q,$$

namely by

$$\frac{\partial Q^{ag}}{\partial \overline{\theta}} = \frac{\partial N^a_{ag}}{\partial \overline{\theta}} \cdot (1 - q) = \frac{N}{\overline{\theta}^2} \cdot (v - p) > 0.$$

[3] García-Gallego and Georgantzís (2009, 2011) and Brécard and Chiroleu-Assouline (2018) investigate a variety of possible changes in the θ distribution that awareness campaigns might induce. Because all three papers adopt the benefits approach, however, they do not notice the very different welfare implications that would emerge from an avoided-damage approach.

Lastly, in a market where all three consumer choices—abstaining, buying green, and buying brown—coexist, we have $N^a_{agb} = n^{ag}$ abstainers, $N^g_{agb} = n^{gb} - n^{ag}$ green buyers, and $N^b_{agb} = N - n^{gb}$ brown buyers. In the benchmark case, the campaign will increase the number of abstainers by more than it reduces the number of brown consumers:

$$\frac{dN^a_{agb}}{d\overline{\theta}} = \frac{dn^{ag}}{d\overline{\theta}} = \frac{N}{\overline{\theta}^2} \cdot \frac{v-p}{1-q},$$

but

$$\frac{dN^b_{agb}}{d\overline{\theta}} = \frac{d(N-n^{gb})}{d\overline{\theta}} = -\frac{N}{\overline{\theta}^2} \cdot \frac{p}{q}.$$

Paradoxically, the number of green consumers therefore falls:

$$\frac{dN^g_{agb}}{d\overline{\theta}} = \frac{dn^{gb} - n^{ag}}{d\overline{\theta}} = -\frac{N}{\overline{\theta}^2} \cdot \left(\frac{v-p}{1-q} - \frac{p}{q}\right) < 0.$$

Because abstention benefits the environment more strongly than green consumption, however, environmental quality, calculated as

$$Q^{agb} = N^a_{agb} \cdot 1 + N^g_{agb} \cdot q + N^b_{agb} \cdot 0,$$

nevertheless improves:

$$\frac{dQ^{agb}}{d\overline{\theta}} = \frac{dN^a_{agb}}{d\overline{\theta}} \cdot 1 + \frac{dN^g_{agb}}{d\overline{\theta}} \cdot q = \frac{N}{\overline{\theta}^2} \cdot \frac{v-p}{1-q} \cdot (1-q) + \frac{N}{\overline{\theta}^2} \cdot \frac{p}{q} \cdot q = \frac{N}{\overline{\theta}^2} \cdot v > 0.$$

It is easy to check that, once again, these effects on consumer behavior and environmental quality are identical if the avoided-damage approach is used to derive them. Crucially, however, the equivalence between the approaches breaks down when it comes to welfare analysis. Writing the integral limits explicitly, and also explicitly as functions of $\overline{\theta}$, we have that welfare in the three market cases using the benefits approach is

$$W^{gb} = \int_0^{n^{gb}(\overline{\theta})} \underbrace{\left[v + \overline{\theta}(1-n/N)q - p\right]}_{U^g} dn + \int_{n^{gb}(\overline{\theta})}^{N} \underbrace{v}_{U^b} dn$$

$$W^{ag} = \int_0^{n^{ag}(\overline{\theta})} \underbrace{\overline{\theta}(1-n/N)}_{U^a} dn + \int_{n^{ag}(\overline{\theta})}^{N} \underbrace{\left[v + \overline{\theta}(1-n/N)q - p\right]}_{U^g} dn$$

$$W^{agb} = \int_0^{n^{ag}(\overline{\theta})} \underbrace{\overline{\theta}(1-n/N)}_{U^a} dn + \int_{n^{ag}(\overline{\theta})}^{n^{gb}(\overline{\theta})} \underbrace{\left[v + \overline{\theta}(1-n/N)q - p\right]}_{U^g} dn + \int_{n^{gb}(\overline{\theta})}^{N} \underbrace{v}_{U^b} dn.$$

Differentiating these expressions using Leibniz' Rule gives

$$\frac{\partial W^{gb}}{\partial \overline{\theta}} = \int_0^{n^{gb}(\overline{\theta})} (1-n/N)q \, dn > 0$$

$$\frac{\partial W^{ag}}{\partial \overline{\theta}} = \int_0^{n^{ag}(\overline{\theta})} (1-n/N) \, dn + \int_{n^{ag}(\overline{\theta})}^{N} (1-n/N)q \, dn > 0$$

$$\frac{\partial W^{agb}}{\partial \bar{\theta}} = \int_{0}^{n^{ag}(\bar{\theta})} (1 - n/N)\, dn + \int_{n^{ag}(\bar{\theta})}^{n^{gb}(\bar{\theta})} (1 - n/N) q\, dn > 0.$$

That is, the awareness campaign unambiguously improves private welfare in all three market cases, reinforcing the positive effect on environmental quality and thereby public welfare.

Under the avoided-damage approach, however, the corresponding welfare expression are

$$\widetilde{W}^{gb} = \int_{0}^{n^{gb}(\bar{\theta})} \underbrace{\left[v - \bar{\theta}(1 - n/N)d - p\right]}_{\widetilde{U}^{g}} dn + \int_{n^{gb}(\bar{\theta})}^{N} \underbrace{v - \bar{\theta}(1 - n/N)}_{\widetilde{U}^{b}} dn$$

$$\widetilde{W}^{ag} = \int_{0}^{n^{ag}(\bar{\theta})} \underbrace{0}_{\widetilde{U}^{a}} dn + \int_{n^{ag}(\bar{\theta})}^{N} \underbrace{\left[v - \bar{\theta}(1 - n/N)d - p\right]}_{\widetilde{U}^{g}} dn$$

$$\widetilde{W}^{agb} = \int_{0}^{n^{ag}(\bar{\theta})} \underbrace{0}_{\widetilde{U}^{a}} dn + \int_{n^{ag}(\bar{\theta})}^{n^{gb}(\bar{\theta})} \underbrace{\left[v - \bar{\theta}(1 - n/N)d - p\right]}_{\widetilde{U}^{g}} dn + \int_{n^{gb}(\bar{\theta})}^{N} \underbrace{v - \bar{\theta}(1 - n/N)}_{\widetilde{U}^{b}} dn,$$

and using Leibniz' Rule gives

$$\frac{\partial \widetilde{W}^{gb}}{\partial \bar{\theta}} = -\int_{0}^{n^{gb}(\bar{\theta})} (1 - n/N) d\, dn - \int_{n^{gb}(\bar{\theta})}^{N} (1 - n/N)\, dn < 0$$

$$\frac{\partial \widetilde{W}^{ag}}{\partial \bar{\theta}} = -\int_{n^{ag}(\bar{\theta})}^{N} (1 - n/N) d\, dn < 0$$

$$\frac{\partial \widetilde{W}^{agb}}{\partial \bar{\theta}} = -\int_{n^{ag}(\bar{\theta})}^{n^{gb}(\bar{\theta})} (1 - n/N) d\, dn - \int_{n^{gb}(\bar{\theta})}^{N} (1 - n/N)\, dn < 0.$$

That is, the awareness campaign unambiguously *reduces* private welfare in all three market cases, offsetting the positive effect on public welfare, and possibly reducing welfare overall.

As noted in the introduction, an informal argument about moral behavior introduced by Miller and Monin (2016) suggests a way to make sense of these seemingly incompatible findings. Specifically, the argument can be used to interpret the benefit approach as modeling consumers who treat buying green or abstaining as moral *opportunities* relative to a norm of buying brown. The resulting utility gain θq or θ then represents the warm glow they feel from availing themselves of those opportunities, and an awareness campaign that raises their utility weight θ increases that warm glow. In contrast, the avoided-damage approach can be thought of as modeling consumers who treat buying green or brown as moral *tests* relative to a norm of abstaining. The resulting utility loss $-\theta d$ or $-\theta$ that they feel represents the cold prickle or guilt they feel from failing those tests, and an awareness campaign that raises their utility weight θ increases that guilt.

Miller and Monin emphasize, moreover, that moral frames vary across individuals, as well as across time for any given individual, depending on the context. This then implies that the true welfare effect of an awareness campaign may be either positive or negative, depending on the mix of consumers that treats either brown consumption or abstention as the norm at any given time.

To capture these ideas it will be useful to slightly change our notation. Utility under the benefit approach, which we have thus far denoted U^x, for $x \in \{a, g, b\}$, is more appropriately denoted $U^{x|b}$, to emphasize the implicit baseline of brown consumption relative to which it is measured. Similarly, utility under the avoided-damage approach, which we have thus far denoted $\widetilde{U}^x$, is more appropriately denoted $U^{x|a}$, to emphasize its implicit baseline of abstention.

Figure 1 compares the utilities under the two approaches, for a case where $\overline{\theta} > v$, so that initially a mix of abstainers and brown consumers exist, and where also $qv > p$, so that the green good, when introduced, captures a positive market share. Notice that the crossing points n^{ag}, n^{ab}, and n^{gb} are identical under the two approaches, as are therefore the equilibrium numbers N^a, N^g, and N^b of abstainers, green consumers, and brown consumers.

Under the benefit approach, however, shown in panel (a), an awareness campaign that increases $\overline{\theta}$ has the effect of pivoting the $U^{a|b}$ curve upwards around its intercept $(N, 0)$ on the right, and pivoting the $U^{g|b}$ curve upwards around its intercept $(N, v - p)$. Both pivots have the effect of increasing the warm glow areas, and thereby improving welfare. In contrast, under the avoided-damage approach shown in panel (b), the awareness campaign pivots the $U^{b|a}$ curve *downwards* around its intercept (N, v) on the right, and pivots the $U^{g|a}$ curve downwards around its intercept $(N, v - p)$. Both pivots have the effect of increasing the guilt areas, and thereby reducing welfare.

One can imagine also that, especially in a market where a green good has captured a large market share, *green* consumption might eventually become the norm. If so, then consumers who continue to abstain should feel warm glow from exceeding that new norm, while consumers who continue to buy brown should feel guilt from falling short of it. Specifically, the abstainers' warm glow should equal $\theta(1 - q)$, from providing benefit 1 rather than the norm of q, while the brown buyers' guilt should equal $\theta(1 - d)$, from causing damages 1 rather than the norm of d.

Figure 2 shows utilities under this third, intermediate possibility, using the same parameters as Figure 1, and thereby again yielding the same crossing points.

Mathematically, we end up with the following expressions for utility under the three norms:

$$\begin{array}{lclcl} U^{a|a} = 0 & & U^{a|g} = \theta(1-q) & & U^{a|b} = \theta \\ U^{g|a} = v - \theta(1-q) - p & \text{vs.} & U^{g|g} = v - p & \text{vs.} & U^{g|b} = v + \theta q - p \\ U^{b|a} = v - \theta & & U^{b|g} = v - \theta q & & U^{b|b} = v. \end{array}$$

Note that for all three choices $x \in \{a, g, b\}$, we have $U^{x|g} = U^{x|a} + \theta(1 - q)$ and $U^{x|b} = U^{x|a} + \theta$.

—

Once the possibility of different utility reference points or norms is allowed for, two key questions arise. The first, already hinted at above, is whether the norm that consumers apply might change over time. Might for example, in a market where initially all consumers buy brown, making that the natural norm, the introduction of a green good that captures a sufficiently large market

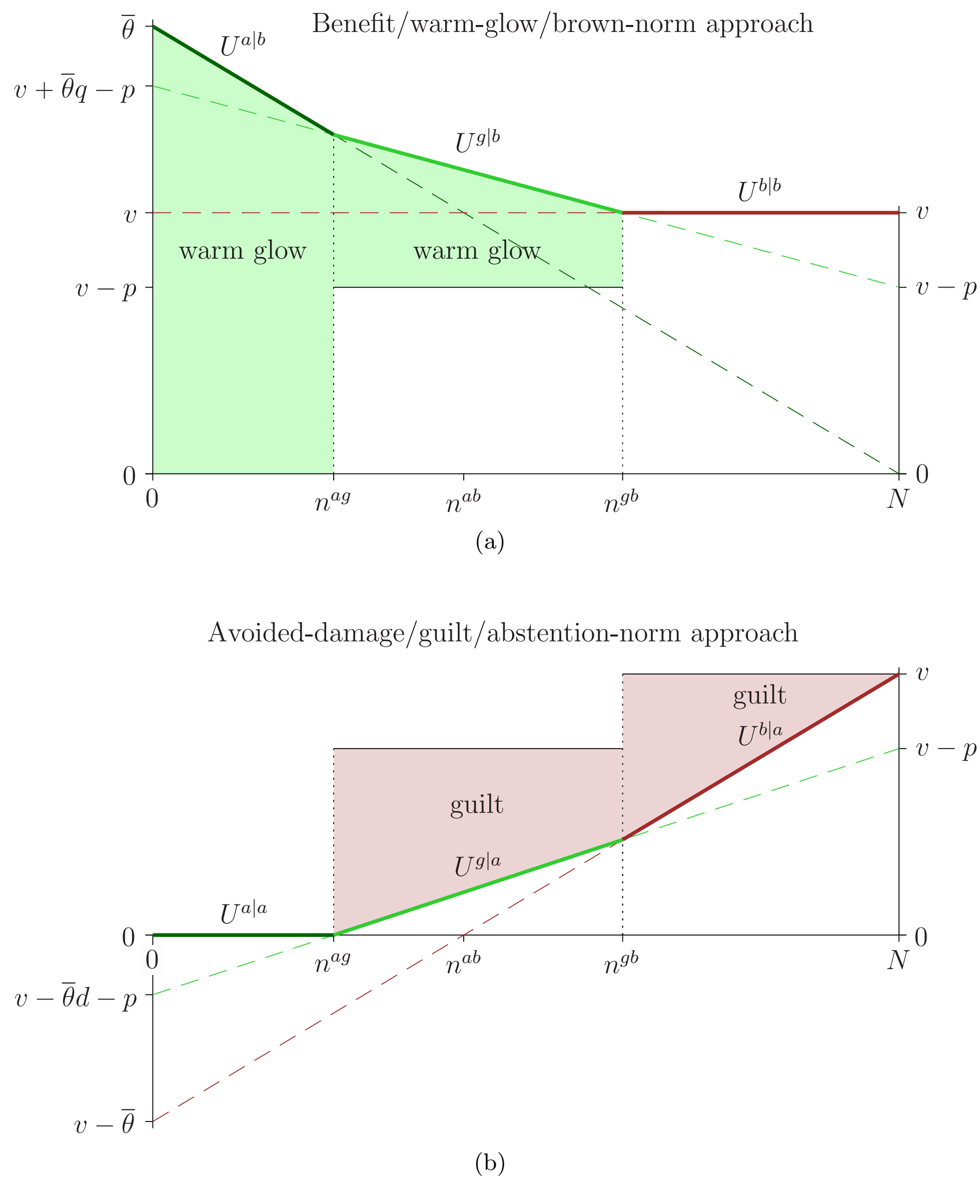

FIGURE 1. Utilities under the benefit and avoided-damages approaches. Under the benefit approach in panel (a), abstainers and green consumers feel warm glow relative to the norm of brown consumption. Under the avoided-damage approach in panel (b), , green and brown consumers feel guilt relative to the norm of abstention.

share eventually change the norm to buying green? The second is whether, if the norm does change, it necessarily does so for all consumers in the market. Should, in order to count as a true norm, *all* consumers in a market agree on what behavior counts as the reference point for warm glow or guilt? Or might each consumer eventually come to see her own behavior as the norm, so that in the above example only consumers who switched to buying green would start adopting the green norm?

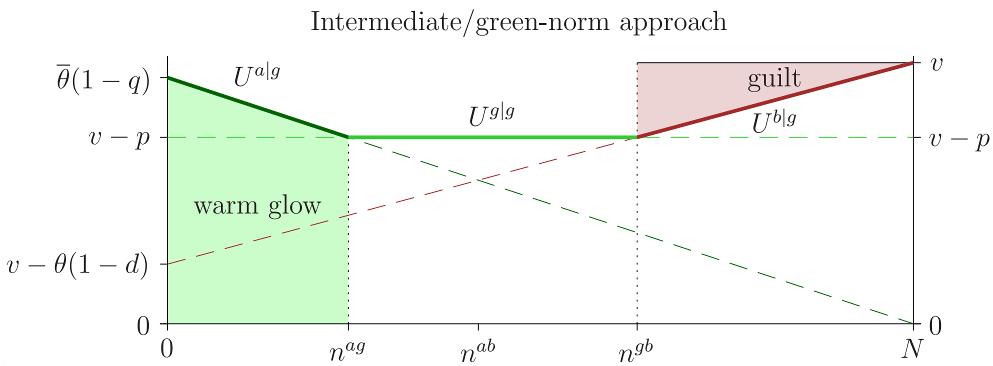

FIGURE 2. Utilities if green consumption is the norm, causing abstainers to feel warm glow, but brown consumers to feel guilt.

Both possibilities—fully universal norms or fully individual, self-centered norms—may seem plausible at first blush, but do not survive closer scrutiny. Consider, for example, a market in which consumers can choose from a range of green goods, each certified to a different standard q or d.[4] If all varieties end up having roughly equal market shares, which is not impossible given the right distribution of preferences and price premia, then it seems odd to assume that just one of the varieties would be viewed by all consumers as the norm relative to which they should feel warm glow or guilt. At the other extreme, if the market shares end up being highly skewed—with one variety dominating, say, and the other ones each occupying a miniscule niche—then it seem odd to assume that niche consumers would consider their own variety the norm.

A more plausible intermediate possibility is that consumers essentially have "mixed feelings": they feel warm glow when comparing their own behavior to that of other consumers who act less green, but guilt when comparing their behavior to that of consumers who act greener. If we make the additional, also plausible assumption that consumers compare themselves more frequently to consumers they encounter more frequently, then we end up with the following, frequency-weighted utilities:

$$U^a = \omega^a U^{a|a} + \omega^g U^{a|g} + \omega^b U^{a|b}$$
$$U^g = \omega^a U^{g|a} + \omega^g U^{g|g} + \omega^b U^{g|b}$$
$$U^b = \omega^a U^{b|a} + \omega^g U^{b|g} + \omega^b U^{b|b},$$

where the weights ω^x on each reference behavior correspond to the relative frequencies N^x/N of each behavior in the market. Note the implication that if an action x is close to universal, so $\omega^x \approx 1$, it reasonably becomes the universal norm, relative to which all consumers compare their actions. Conversely, if an action x is eccentric, so $\omega^x \approx 0$, most consumers ignore it.

A crucial final assumption is that when a *new* action enters the picture—a green good is introduced, say—consumers will initially evaluate it according to the *current* distribution of actions, and only later, once a new distribution has been established, evaluate it according to that new distribution. Consumer expectations are adaptive, in other words, not forward-looking.

[4] This could be because of competing eco-labels or because of eco-label gradation. See Li and van 't Veld (2015).

The next section analyzes how, under these assumptions, introducing a green good affects welfare in both the short and long run.

4. Norm effects on welfare

Consider once again the three initial market settings of Section 2.

4.1. Initial all-brown market

In an initial all-brown market, utility equals $U^{b|b}$ for all consumers, so welfare is

$$W_b^b = \int_{\mathcal{N}} \underbrace{v}_{U^{b|b}} \, dn = Nv.$$

If now a green good is introduced, all consumers will evaluate the utility from it as $U^{g|b}$, i.e., expect it to yield warm glow $\theta(n)q$ from providing benefit q relative to the current norm of brown consumption, in exchange for the price premium p. As was shown in Section 2.1, a total of N_{gb}^g consumers, with weights $\theta(n) > p/q = \theta^{gb}$, will then switch, causing welfare to increase initially to

$$W_b^{gb} = \int_{\mathcal{N}_{gb}^g} \underbrace{[v + \theta(n)q - p]}_{U^{g|b}} \, dn + \int_{\mathcal{N}_{gb}^b} \underbrace{v}_{U^{b|b}} \, dn$$

and giving rise to welfare improvement

$$W_b^{gb} - W_b^b = \int_{\mathcal{N}_{gb}^g} \underbrace{[\theta(n)q - p]}_{U^{g|b} - U^{b|b} > 0} \, dn > 0$$

from the green consumers' increased surplus.

Eventually, however, utility from buying the two goods will adjust to the new distribution of actions, changing to

$$\begin{aligned} U_{gb}^g &= \omega_{gb}^g U^{g|g} + \omega_{gb}^b U^{g|b} \\ U_{gb}^b &= \omega_{gb}^g U^{b|g} + \omega_{gb}^b U^{b|b}, \end{aligned}$$

where $\omega_{gb}^g = N_{gb}^g/N$ and $\omega_{gb}^b = N_{gb}^b/N = 1 - \omega_{gb}^g$.

The utility change in turn causes welfare to change to

$$\begin{aligned} W_{gb}^{gb} &= \int_{\mathcal{N}_{gb}^g} \Big\{ \omega_{gb}^g \underbrace{[v - p]}_{U^{g|g}} + \omega_{gb}^b \underbrace{[v + \theta(n)q - p]}_{U^{g|b}} \Big\} \, dn + \int_{\mathcal{N}_{gb}^b} \Big\{ \omega_{gb}^g \underbrace{[v - \theta(n)q]}_{U^{b|g}} + \omega_{gb}^b \underbrace{v}_{U^{b|b}} \Big\} \, dn \\ &= W_b^{gb} - \int_{\mathcal{N}} \omega_{gb}^g \theta(n) q \, dn \end{aligned} \tag{5}$$

That is, the effect of the norm change is to reduce *all* consumers' utility by $\omega_{gb}^g \theta(n) q$.

Underlying this are two separate effects. First, consumers who switched to buying green continue to feel warm glow w.r.t. the fraction ω_{gb}^b of brown consumers, but will eventually no longer feel warm glow w.r.t. the fraction ω_{gb}^g of their fellow green consumers. As a result, they lose utility

$-\omega_{gb}^g\theta(q)$. Second, consumers who stuck with buying brown continue to feel no guilt w.r.t. the fraction ω_{gb}^b of their fellow brown consumers, but will eventually start feeling guilt w.r.t. the fraction ω_{gb}^g of green consumers. As a result, they lose utility $-\omega_{gb}^g\theta(q)$ also.

Importantly, once the norm has changed, there is no "going back": because utility falls by $\omega_{gb}^g\theta q$ for *both* types of consumers, both will continue to prefer whatever action they chose initially.

The overall, long-run effect on welfare is

$$W_{gb}^{gb} - W_b^b = \int_{\mathcal{N}_{gb}^g} \underbrace{[\theta(n)q - p]}_{+} \, dn - \int_{\mathcal{N}} \underbrace{\omega_{gb}^g \theta(n) q}_{+} \, dn \gtreqless 0,$$

which in general is ambiguous in sign.

In the benchmark case with linear weights, however, the initial consumer-surplus effect evaluates to

$$\int_{\mathcal{N}_{gb}^g} [\theta(n)q - p] \, dn = \tfrac{1}{2} N \frac{(\overline{\theta}q - p)^2}{\overline{\theta}q},$$

and the eventual norm effect to

$$-\int_{\mathcal{N}} \omega_{gb}^g \theta(n) q = -\tfrac{1}{2} N(\overline{\theta}q - p). \tag{6}$$

The net effect is therefore unambiguously negative:

$$W_{gb}^{gb} - W_b^b = -\tfrac{1}{2} N(\overline{\theta}q - p)\frac{p}{\overline{\theta}q} < 0.$$

In other words, regardless of the green good's quality q or price premium p, the initial increase in warm glow that draws some consumers to buying green ends up being swamped by eventual reductions in warm glow and increases in guilt as norms of green behavior shift.

The surprising implication, explored further in Section 5 below, is that introducing a green good may be welfare reducing overall, unless it raises environmental quality Q and thereby public welfare BQ sufficiently. Recall from equation (1) in Section 2.1 that environmental quality increases by

$$\Delta Q = Q^{gb} - Q^b = N_{gb}^g \cdot q$$

as a result of the green-good introduction. In the benchmark case, this evaluates to

$$\Delta Q = N(\overline{\theta}q - p)\frac{1}{\overline{\theta}} > 0.$$

Comparing this to expression (6) for the norm effect shows that the positive effect $B\Delta Q$ on public welfare exactly offsets that negative norm effect, leaving only the positive consumer-surplus effect, if the welfare weight on environmental benefits equals $B = \tfrac{1}{2}\overline{\theta}$. This critical value, too, will come back in Section 5.

4.2. **Initial all-abstention market**

In a setting where initially no good exists, utility equals $U^{a|a}$ for all consumers, so welfare is

$$W_a^a = \int_{\mathcal{N}} \underbrace{0}_{U^{a|a}} \, dn = 0.$$

If now a green good is introduced, all consumers will evaluate the utility of it as $U^{g|a}$, i.e., expect it to give rise to guilt $\theta(n)(1-q)$ because of the damages $d = 1-q$ associated with green consumption, offset by net use value $v - p$. As was shown in Section 2.2, a total of N^g_{ag} consumers, with weights $\theta(n) < (v-p)/(1-q) = \theta^{ag}$, will choose to buy the green good, causing welfare to increase initially to

$$W^{ag}_a = \int_{\mathcal{N}^a_{ag}} \underbrace{0}_{U^{a|a}} \, dn + \int_{\mathcal{N}^g_{ag}} \underbrace{[v - \theta(n)(1-q) - p]}_{U^{g|a}} \, dn,$$

and giving rise to welfare improvement

$$W^{ag}_a - W^a_a = \int_{\mathcal{N}^g_{ag}} \underbrace{[v - \theta(n)(1-q) - p]}_{U^{g|a} - U^{a|a} > 0} \, dn > 0$$

from the green consumers' surplus.

As norms of behavior adjust, however, utility eventually changes to

$$U^a = \omega^a_{ag} U^{a|a} + \omega^g_{ag} U^{a|g}$$
$$U^g = \omega^a_{ag} U^{g|a} + \omega^g_{ag} U^{g|g},$$

where $\omega^a_{ag} = N^a_{ag}/N$ and $\omega^g_{ag} = N^g_{ag}/N = 1 - \omega^a_{ag}$. As a result, welfare changes to

$$\begin{aligned} W^{ag}_{ag} = &\int_{\mathcal{N}^a_{ag}} \Big\{ \omega^a_{ag} \underbrace{0}_{U^{a|a}} \, dn + \omega^g_{ag} \underbrace{\theta(n)(1-q)}_{U^{a|g}} \Big\} \, dn \\ &+ \int_{\mathcal{N}^g_{ag}} \Big\{ \omega^a_{ag} \underbrace{[v - \theta(n)(1-q) - p]}_{U^{g|a}} + \omega^g_{ag} \underbrace{[v-p]}_{U^{g|g}} \Big\} \, dn \\ = &\, W^{ag}_a + \int_{\mathcal{N}} \omega^g_{ag} \theta(n)(1-q) \, dn. \end{aligned} \tag{7}$$

That is, the effect of the norm change is to increase all consumers' utility by $\omega^g_{ag}\theta(n)(1-q)$.

The additional gain arises for two reasons. First, consumers who stuck with abstention start feeling warm glow $\theta(n)(1-q)$ w.r.t. the fraction ω^g_{ag} of consumers who switched to green. Second, consumers who switched to green stop feeling guilt $\theta(n)(1-q)$ w.r.t. to the fraction ω^g_{ag} of consumers who, like them, also switched to green.

The overall, long-run effect on welfare is unambiguously positive,

$$W^{ag}_{ag} - W^a_a = \int_{\mathcal{N}^g_{ag}} \underbrace{[v - \theta(n)(1-q) - p]}_{+} \, dn + \int_{\mathcal{N}} \underbrace{\omega^g_{ag}\theta(n)(1-q)}_{+} \, dn > 0,$$

with the eventual norm effect reinforcing the initial consumer-surplus effect.

In the benchmark case with linear weights, the consumer-surplus effect evaluates to

$$\int_{\mathcal{N}^g_{ag}} [v - \theta(n)(1-q) - p] \, dn = \tfrac{1}{2} N \frac{(v-p)^2}{\bar{\theta}(1-q)}$$

and the norm effect to

$$\int_{\mathcal{N}} \omega^g_{ag}\theta(n)(1-q)\,dn = \tfrac{1}{2}N(v-p). \tag{8}$$

Since we must have

$$\frac{v-p}{\overline{\theta}(1-q)} = \frac{\theta^{ag}}{\overline{\theta}} < 1$$

for the green good to not completely take over the market, it follows again that the norm effect exceeds the consumer-surplus effect in magnitude.

Recall also from equation (2) in Section 2.2 that environmental quality decreases by

$$\Delta Q = Q^{ag} - Q^{a} = -N^g_{ag}\cdot(1-q),$$

as a result of the green-good introduction. In the benchmark case, this evaluates to

$$\Delta Q = -N(v-p)\frac{1}{\overline{\theta}}.$$

Comparing this to the norm effect (8) shows that the negative effect $B\Delta Q$ on public welfare exactly offsets the positive norm effect, leaving only the positive consumer-surplus effect, if the welfare weight on environmental benefits equals $B = \frac{1}{2}\overline{\theta}$—the same critical value as found in the previous subsection.

4.3. **Initial mixed market**

In an initial mixed market, where the norm is a mix of abstention and brown consumption, utility equals $U^a_{ab} = \omega^a_{ab}U^{a|a} + \omega^b_{ab}U^{a|b}$ for abstainers, and $U^b_{ab} = \omega^a_{ab}U^{b|a} + \omega^b_{ab}U^{b|b}$ for brown consumers. Welfare is therefore

$$\begin{aligned} W^{ab}_{ab} &= \int_{\mathcal{N}^a_{ab}} U^a_{ab}\,dn + \int_{\mathcal{N}^b_{ab}} U^b_{ab}\,dn \\ &= \int_{\mathcal{N}^a_{ab}} \Big\{\omega^a_{ab}\underbrace{0}_{U^{a|a}} + \omega^b_{ab}\underbrace{\theta(n)}_{U^{a|b}}\Big\}\,dn + \int_{\mathcal{N}^b_{ab}} \Big\{\omega^a_{ab}\underbrace{[v-\theta(n)]}_{U^{b|a}} + \omega^b_{ab}\underbrace{v}_{U^{b|b}}\Big\}\,dn, \end{aligned}$$

where $\omega^a_{ab} = N^a_{ab}/N$ and $\omega^b_{ab} = N^b_{ab}/N$.

If now a green good is introduced, consumers will initially evaluate the utility of it according to the current norm. Specifically, they will expect it to give rise to guilt-tinged surplus $v-\theta(n)(1-q)-p$ w.r.t. abstainers, but warm-glow-tinged surplus $v+\theta(n)q-p$ w.r.t. brown consumers. As was shown in Section 2.3, a total of N^g_{agb} consumers, with weights $\theta(n) \in [(v-p)/(1-q), p/q] = [\theta^{ag}, \theta^{gb}]$, will choose to buy the green good, causing welfare to increase initially to

$$\begin{aligned} W^{agb}_{ab} &= \int_{\mathcal{N}^a_{agb}} U^a_{ab}\,dn + \int_{\mathcal{N}^g_{agb}} U^g_{ab}\,dn + \int_{\mathcal{N}^a_{agb}} U^b_{ab}\,dn \\ &= \int_{\mathcal{N}^a_{agb}} \Big\{\omega^a_{ab}\underbrace{0}_{U^{a|a}} + \omega^b_{ab}\underbrace{\theta(n)}_{U^{a|b}}\Big\}\,dn \\ &\quad + \int_{\mathcal{N}^g_{agb}} \Big\{\omega^a_{ab}\underbrace{[v-\theta(n)(1-q)-p]}_{U^{g|a}} + \omega^b_{ab}\underbrace{[v+\theta(n)q-p]}_{U^{g|b}}\Big\}\,dn \end{aligned}$$

$$+ \int_{\mathcal{N}_{agb}^{b}} \Big\{ \omega_{ab}^{a} \underbrace{[v - \theta(n)]}_{U^{b|a}} + \omega_{ab}^{b} \underbrace{v}_{U^{b|b}} \Big\} \, dn,$$

for a welfare gain of

$$W_{ab}^{agb} - W_{ab}^{ab} = \int_{\mathcal{N}_{agb}^{a \to g}} \underbrace{[v - \theta(n)(1-q) - p]}_{U_{ab}^{g} - U_{ab}^{a} > 0} \, dn + \int_{\mathcal{N}_{agb}^{b \to g}} \underbrace{[\theta(n)q - p]}_{U_{ab}^{g} - U_{ab}^{b} > 0} \, dn.$$

But then norms adjust, and utility changes to

$$\begin{aligned} U^{a} &= \omega_{agb}^{a} U^{a|a} + \omega_{agb}^{g} U^{a|g} + \omega_{agb}^{b} U^{a|b} \\ U^{g} &= \omega_{agb}^{a} U^{g|a} + \omega_{agb}^{g} U^{g|g} + \omega_{agb}^{b} U^{g|b} \\ U^{b} &= \omega_{agb}^{a} U^{b|a} + \omega_{agb}^{g} U^{b|g} + \omega_{agb}^{b} U^{b|b}, \end{aligned}$$

where $\omega_{agb}^{a} = N_{agb}^{a}/N$, $\omega_{agb}^{g} = N_{agb}^{g}/N$, and $\omega_{ab}^{b} = N_{agb}^{b}/N$. As a result, welfare changes to

$$\begin{aligned} W_{agb}^{agb} = & \int_{\mathcal{N}_{agb}^{a}} \Big\{ \omega_{agb}^{a} \underbrace{0}_{U^{a|a}} + \omega_{agb}^{g} \underbrace{\theta(n)(1-q)}_{U^{a|g}} + \omega_{agb}^{b} \underbrace{\theta(n)}_{U^{a|b}} \Big\} \, dn \\ & + \int_{\mathcal{N}_{agb}^{g}} \Big\{ \omega_{agb}^{a} \underbrace{[v - \theta(n)(1-q) - p]}_{U^{g|a}} + \omega_{agb}^{g} \underbrace{[v - p]}_{U^{g|g}} + \omega_{agb}^{b} \underbrace{[v + \theta(n)q - p]}_{U^{g|b}} \Big\} \, dn \\ & + \int_{\mathcal{N}_{agb}^{b}} \Big\{ \omega_{agb}^{a} \underbrace{[v - \theta(n)]}_{U^{b|a}} + \omega_{agb}^{g} \underbrace{[v - \theta(n)q]}_{U^{b|g}} + \omega_{agb}^{b} \underbrace{v}_{U^{b|b}} \Big\} \, dn. \end{aligned}$$

Importantly, because some green consumers switched from abstention, while some switched from buying brown, the term ω_{agb}^{g} can be rewritten as the sum of the weights on both subgroups, i.e., as $(\omega_{ab}^{a} - \omega_{agb}^{a}) + (\omega_{ab}^{b} - \omega_{agb}^{b})$. Using this, the new welfare expression can be written as

$$W_{agb}^{agb} = W_{ab}^{agb} + \int_{\mathcal{N}} \Big\{ (\omega_{ab}^{a} - \omega_{agb}^{a}) \theta(n)(1-q) - (\omega_{ab}^{b} - \omega_{agb}^{b}) \theta(n) q \Big\} \, dn.$$

Comparing this to earlier expressions (5) and (7) shows that the norm effect for the initial mixed market is a weighted sum of the norm effect $\int_{\mathcal{N}} \omega_{ag}^{g} \theta(n)(1-q)$ for the initial all-abstention market and the norm effect $-\int_{\mathcal{N}} \omega_{gb}^{g} \theta(n) q \, dn$ for the initial all-brown market.

Specifically, the $N_{agb}^{a \to g} = N_{ab}^{a} - N_{agb}^{a}$ consumers who switch from abstaining to buying green do so anticipating a utility gain of $v - \theta(n)(1-q) - p$. In long-run equilibrium, however, all N consumers end up adding glow $(\omega_{ab}^{a} - \omega_{agb}^{a})\theta(n)(1-q)$ or reducing guilt by that same amount w.r.t. this subgroup of switchers, including the switchers themselves. At the same time, the $N_{agb}^{b \to g} = N_{ab}^{b} - N_{agb}^{b}$ consumers who switch from buying brown to buying green do so anticipating a utility gain of $\theta(n)q - p$. In long-run equilibrium, however, all N consumers end up losing glow $(\omega_{ab}^{b} - \omega_{agb}^{b})\theta(n)q$ or increasing guilt by that same amount w.r.t. this second subgroup of switchers, again including the switchers themselves.

Once again, how the initial consumer-surplus increase compares to the eventual norm changes depends on parameters, and in particular on the distribution of utility weights.

In the benchmark linear case, using that the index n^{ab} of a consumer who is initially just indifferent between buying green or brown is given by

$$\overline{\theta}(1 - n^{ab}/N) = \theta^{ab}$$
$$\Leftrightarrow \quad n^{ab} = N\left(1 - \frac{\theta^{ab}}{\overline{\theta}}\right) = N\left(1 - \frac{v}{\overline{\theta}}\right),$$

the consumer-surplus effect evaluates to

$$\int\limits_{\mathcal{N}_{agb}^{a\to g}} \big[v - \theta(n)(1-q) - p\big]\, dn + \int\limits_{\mathcal{N}_{agb}^{b\to g}} \big[\theta(n)q - p\big]\, dn = \tfrac{1}{2}N\frac{(vq-p)^2}{\overline{\theta}(1-q)q} > 0,$$

and the norm effect to

$$\int\limits_{\mathcal{N}} \Big\{(\omega_{ab}^{a} - \omega_{agb}^{a})\theta(n)(1-q) - (\omega_{ab}^{b} - \omega_{agb}^{b})\theta(n)q\Big\}\, dn = 0!$$

Surprisingly, the added glow and reduced guilt induced by switchers from abstention turns out to in aggregate exactly cancel out the lost glow and increased guilt induced by switchers from brown consumption. In the benchmark case, in other words, introducing a green good into a non-covered market unambiguously improves private welfare, because there is no long-run negative norm effect to offset the initial positive consumer-surplus effect.

Surprisingly also, the green-good introduction has no effect on environmental quality and thereby public welfare. Recall from (4) that environmental quality changes by

$$\Delta Q = Q^{agb} - Q^{ab} = -N_{agb}^{a\to g} \cdot (1-q) + N_{agb}^{b\to g} \cdot q \gtreqless 0, \tag{9}$$

which in general is ambiguous. In the benchmark case, however, this evaluates to

$$\Delta Q = 0.$$

The reason is that in expression (9) the ratio of the number of switchers from abstention to buying green, $N_{agb}^{a\to g}$, to the number of switchers from buying brown to green, $N_{agb}^{b\to g}$, turns out reduce to $q/(1-q)$, thereby exactly counterbalancing the ratio $-(1-q)/q$ of these two subgroups' per-capita effects on environmental quality.

These zero effects in the benchmark are of course knife-edge, depending very much on the strong linearity assumptions of the model. Moreover, they hold only if the green good's quality-price tradeoff is not so favorable that *all* initial abstainers switch: at least some consumers must continue to abstain. Nevertheless, the stark difference between the findings for this scenario and those for the covered-market scenario of Section 4.1 point to the importance of accounting for abstainers when analyzing green markets.

5. Norm effect on optimal performance standards

The analysis thus far has treated the green good's quality q and price premium p as arbitrary, requiring only that their ratio is not so high as to drive all abstainers out of the market, nor so low as to attract no buyers. As noted in the introduction, however, the actual quality of green goods is typically set by eco-label programs, and the price premium is driven by the additional production cost of providing that quality. This sets up a tradeoff for the programs: raising per-unit quality and thereby price will tend to reduce the number of units sold. Moreover, different eco-label sponsors—industry, government, or environmental NGOs—will respond to this tradeoff differently, because of their different objectives.

Van 't Veld and Kotchen (2011), Manasakis et al. (2013), and Li and van 't Veld (2015) specifically find that a government sponsor will set a more stringent quality standard than an industry sponsor, because the government will take account of consumer surplus and environmental benefits on top of producer surplus. The analysis of all three papers adopts the standard benefit approach, however, with fixed utility weights. Implicitly, therefore, all three papers assume that brown consumption is the norm relative to which green consumers experience warm glow, and that this norm never changes. Moreover, all three papers assume that market is covered, so abstention plays no role. This is the setting of Section 4.1, where it was shown that the norm effect of introducing a green good is negative, and potentially so strong as to outweigh any positive consumer-surplus effect. How might a government sponsor behave differently if it takes this norm effect into account, i.e., allows for the fact that the initial brown norm might change over time?

To investigate this, I borrow the framework used by Li and van 't Veld (2015), including their assumption that green firms incur a firm-specific fixed cost s over and above the additional variable cost of green production, as well as a certification fee k. A minor difference is that Li and van 't Veld assume utility to be concave and firms' variable production cost linear in environmental benefits. Here, because utility is assumed linear in q, the variable cost $c(q)$ must be assumed convex in order to ensure interior solutions.[5]

Assume each firm produces one unit of the good, and treat s as an increasing function of an index variable x on $[0, N]$; This makes it easy to think of ordering firms by increasing s, by lining them up in order of that index variable. The first firm, with index $x = 0$, will have the lowest fixed costs $s(0)$, while firms with higher x will have higher fixed costs.

With this specification, firms will enter a market in which an eco-label certifier operates and sets a standard of q until the last entrant makes zero profits, so

$$\pi = p - c(q) - s(n^s) - k = 0, \tag{10}$$

where n^s denotes green-good supply.

When the green good comes onto the market, consumers will buy it if their utility

$$U^{g|b} = v + \theta(n)q - p$$

from doing so exceeds utility

$$U^{b|b} = v$$

[5] The difference is easily reconciled by distinguishing utility benefit q from the measurable, physical environmental quality e (e.g., emissions reduced, or endangered animals saved) that gives rise to that benefit. This paper follows much of the eco-labeling literature in ignoring this distinction. However, if q is interpreted as the warm glow received by a consumer with utility weight $\theta = 1$, where this warm glow is concave in e, i.e., $q = f(e)$, $f' > 0$, $f'' < 0$, then Li and van 't Veld's linear variable cost αe of achieving a given level of environmental quality e translates into a convex variable costs $c(q) = \alpha f^{-1}(e)$ of achieving a given level of warm glow q.

from buying brown. Demand n^d for the green good will therefore satisfy condition

$$\theta(n^d)q - p = 0. \tag{11}$$

Once the green good has become established, green consumers' utility will change to

$$U^g_{gb} = \omega^g_{gb}\underbrace{[v-p]}_{U^{g|g}} + \omega^b_{gb}\underbrace{[v+\theta(n)q-p]}_{U^{g|b}}$$

and

$$U^b_{gb} = \omega^g_{gb}\underbrace{[v-\theta(n)q]}_{U^{b|g}} + \omega^b_{gb}\underbrace{v}_{U^{b|b}}.$$

This still implies, though, that $U^g_{gb} \geq U^b_{gb}$ for consumers with $\theta(n) \geq \theta(n^d)$, and so will not change demand.

In equilibrium, the market will clear at green-good market share $n^* \equiv n^s = n^d$ given by condition

$$\theta(n)q - c(q) - s(n) - k = 0. \tag{12}$$

Treating this condition as implicitly defining $n^* = n(q)$, we have by the Implicit Function Theorem that

$$n'(q) = \frac{\theta(n(q)) - c'(q)}{-\theta'(n(q))q + s'(n(q))}. \tag{13}$$

Consider now an industry-sponsored eco-label program that aims to maximize the aggregate producer surplus $PS = n^*\pi$ of all firms it certifies. This program faces problem

$$\max_{q,n^*,p} PS = n^*\Big[p - c(q) - \frac{1}{n^*}\int_0^{n^*} s(x)\,dx - k\Big],$$

subject to (10) and (11), where $n^* = n^s = n^d$. Substituting in those constraints simplifies the problem to

$$\max_q PS = n(q)s(n(q)) - \int_0^{n(q)} s(x)\,dx,$$

with first-order condition

$$\frac{dPS}{dq} = n(q)s'(n(q))n'(q) = 0.$$

Since the terms $n(q)$ and $s'(n(q))$ in this condition are both positive, the optimal industry standard q^i will be such that $n'(q^i) = 0$, i.e., the size of the green market is maximized.

How does this standard compare to that of a welfare-maximizing government? Abusing notation by now using W to denote overall welfare (including producer surplus PS and public environmental benefits denoted PB), and also dropping the constant aggregate use-value component Nv of welfare as understood, we can write the government's optimization problem as

$$\max_{q,n^*,p} W = \underbrace{\int_0^{n^*}[\theta(n)q - p]\,dn}_{CS} + \underbrace{\int_0^{n^*}[p - c(q) - s(x) - k]\,dx}_{PS} - \underbrace{\frac{n^*q}{N}\int_0^N \theta(n)\,dn}_{NE} + \underbrace{Bn^*q}_{PB},$$

subject again to conditions (10) and (11). The third term, labeled *NE*, represents the norm effect, which enters negatively. In Section 4.1, the norm effect from introducing a green good into an all-brown market was shown to equal

$$-\int\limits_{\mathcal{N}} \omega_{gb}^{g} \theta(n) q \, dn,$$

where $\omega_{gb}^{g} = N_{gb}^{g}/N$. In the notation of this section, this becomes

$$-\int_0^N \frac{n^*}{N} \theta(n) q \, dn = -\frac{n^* q}{N} \int_0^N \theta(n) \, dn.$$

After substituting in the constraints, the government's problem simplifies to

$$\max_q W = \int_0^{n(q)} [\theta(n)q - c(q) - s(n) - k] \, dn - \frac{n(q)q}{N} \int_0^N \theta(n) \, dn + Bn(q)q,$$

with first-order condition

$$\frac{dW}{dq} = \underbrace{\int_0^{n(q)} \theta(y) \, dy - n(q)c'(q)}_{\frac{d(CS+PS)}{dq}} - \underbrace{\frac{n'(q)q + n(q)}{N} \int_0^N \theta(y) \, dy}_{\frac{dNE}{dq}} + \underbrace{B[n'(q)q + n(q)]}_{\frac{dPB}{dq}} = 0.$$

Evaluated at q^i, where $n'(q^i) = 0$ and therefore, from (13), $\theta(n(q^i)) = c'(q^i)$, the derivative of welfare becomes

$$\left.\frac{dW}{dq}\right|_{q^i} = \underbrace{\int_0^{n(q^i)} \theta(y) \, dy - n(q^i)\theta(n(q^i))}_{\left.\frac{dCS}{dq}\right|_{q^i} > 0} - \underbrace{\frac{n(q^i)}{N} \int_0^N \theta(y) \, dy}_{\left.\frac{dNE}{dq}\right|_{q^i} > 0} + \underbrace{Bn(q^i)}_{\left.\frac{dPB}{dq}\right|_{q^i} > 0}. \tag{14}$$

In the absence of norm effects, as in Li and van 't Veld's (2015) original analysis, this derivative is strictly positive, indicating that a government sponsor will optimally set a standard q^g stricter than the industry standard q^i. Once norm effects enter the picture, however, the derivative may well be negative, indicating that a government sponsor may optimally set a weaker standard than an industry sponsor would.

In fact, the norm effect may be so strong that the government may prefer not to introduce a green good at all. To see this, consider again the benchmark case where $\theta(n)$ is linear, and assume that the switching cost $s(x)$ is linear as well, of the simple form mx for some constant m. Welfare in this case evaluates to

$$W = \underbrace{\tfrac{1}{2}\overline{\theta}\frac{(n(q))^2}{N}q}_{CS} + \underbrace{\tfrac{1}{2}m(n(q))^2}_{PS} - \underbrace{\tfrac{1}{2}\overline{\theta}n(q)q}_{NE} + \underbrace{Bn(q)q}_{PB}. \tag{15}$$

or, rearranging and using that $\theta(n) = \overline{\theta}(1 - n/N)$ and $s(n) = mn$,

$$W = \underbrace{\tfrac{1}{2}n(q)\left[s(n(q)) - \theta(n(q))q\right]}_{CS+PS+NE} + \underbrace{Bn(q)q}_{PB}. \tag{16}$$

We know moreover from equilibrium condition (12) that

$$s(n(q)) - \theta(n(q))q = -c(q) - k < 0,$$

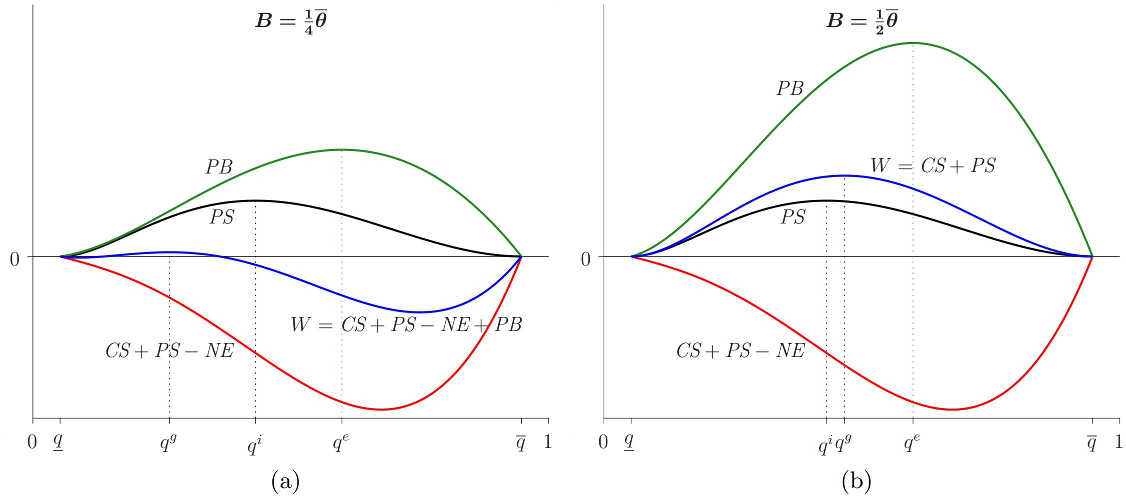

FIGURE 3. Welfare as a function of environmental quality q, and implied optimal standards q^g, q^i, and q^e of government-, industry-, and environmental NGO-sponsored labels. In panel (a), the welfare weight on public benefits is low, $B = \frac{1}{4}\overline{\theta}$, whereas in panel (b) it equals the critical value $B = \frac{1}{2}\overline{\theta}$ where $-NE + PB = 0$.

implying that the term in brackets in (16) is negative. In other words, if the weight B on public environmental benefits is sufficiently small, the total contribution of a green good to welfare—summing consumer surplus, producer surplus, and norm effects—will be negative, and welfare is therefore higher if the market is left all-brown.

Figure 3 illustrates this using a numerical example,[6] plotting the various welfare components against q. For a green market to be viable, the standard must lie in interval $(\underline{q}, \overline{q})$, where the warm glow of at least the greenest consumer is sufficient to cover production costs.

In both panels of the figure, the green curve shows public benefits PB, the black curve shows producer surplus, PS, the red curve shows the overall contribution of the green good to private welfare only, $CS + PS - NE$, and the blue curve shows total welfare, W, including public benefits. In panel (a), the welfare weight applied to public benefits is $B = \frac{1}{4}\overline{\theta}$. At this weight, welfare is just barely positive at low values of the standard, and the welfare-maximizing standard q^g lies well below the PS-maximizing industry standard q^i. (Also shown along the horizontal axis is the standard q^e that a PB-maximizing environmental NGO would choose if it sponsored an eco-label.)

In panel (b), the welfare weight is higher, at $B = \frac{1}{2}\overline{\theta}$. As is evident from expression (15), and consistent with the analysis of Section 4, public benefits at this critical weight exactly offset the negative norm effect, reducing welfare to the sum of consumer and producer surplus only. Moreover, from (14), consumer surplus at q^i is still increasing in q, making it optimal for the government to set a standard q^g stricter than q^i.

[6] The example uses parameters $\overline{\theta} = 1$, $N = 1$, and $k = 0.01$, and functional forms $c(q) = q^2$ and $s(x) = x$.

6. Conclusion

The theoretical literature analyzing economic, environmental, and welfare outcomes of green markets has tended to be cavalier about consumer motivations. Some papers treat green goods as providing environmental benefits that enter green consumers' utility positively; others treat green goods as avoiding environmental damages that enter green consumers' utility negatively. This paper clarifies what set of assumptions is required for these two approaches to be equivalent, and what happens if those assumptions fail to hold.

One such assumption is that consumers' utility weights on either benefits or damages remain fixed. If environmental awareness campaigns raise these utility weights, however, then the benefit approach predicts a welfare increase, whereas the avoided-damage approach predicts a welfare reduction.

While seemingly incompatible, these results make sense if the benefit approach is interpreted as having consumers enjoy warm glow from exceeding a norm of brown consumption, while the damage approach has consumers suffer guilt from falling short of a norm of abstention.

This raises the question where these norms come from, and whether they are necessarily fixed. The main finding of this paper is that, if norms depend in a plausible way on the mix of behaviors in a market, and change over time when that mix changes, a number of counterintuitive implications follow.

First, introducing a green good in a market where initially all consumers are brown may end up reducing long-run welfare overall, once the induced change in norms is taken into account. Second, the change in norms from introducing a green good where initially no good exists at all may turn an initial welfare reduction (because of environmental damages still caused by green-good consumption) into a long-run increase. Third, introducing a green good into a market with a mix of abstainers and brown consumers gives rise to offsetting norm effects. Lastly, when a government eco-label sponsor anticipates norm changes, it may optimally pick a certification standard that is weaker than an industry sponsor would pick, and may even optimally forgo introducing the green good altogether.

Although not the main focus of the paper, the first three results combined highlight that accounting for abstention in green markets can make a qualitative difference to predicted outcomes. This, too, has not been recognized in the theoretical literature: ignoring abstention (i.e., treating the market as covered) has generally been treated as an innocuous simplification.

All four results also raise a novel empirical question, namely how private utility from green goods—whether in terms of warm glow enjoyed or guilt avoided—compares to public utility in terms of magnitude. If successful introduction of a green product increases both sources of utility—as is true by revealed preference in the short run—then knowing their relative magnitude is perhaps not that important. But if norm effects, or changes in abstention, push private and public utility in opposite directions, then the comparison obviously does matter.

A theoretical extension left for future work concerns the comparison of eco-label programs to alternatives such as direct regulation, possibly in the form of mandating green production. Regulating brown production away clearly improves environmental quality and thereby public utility, while at the same time reducing private utility through removing a consumption option. Not clear is how norm effects might alter this tradeoff, particularly if regulations, by removing the element of moral *choice*, crowd out warm glow or guilt.

Most generally, this paper points to the importance of investigating empirically (through lab experiments or otherwise) how green consumption norms are formed, and how they drive feelings of warm glow or cold prickle.

References

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence, *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments, *The Quarterly Journal of Economics* **110**(1): 1–21.

Brécard, D. and Chiroleu-Assouline, M. (2018). Component-free strategy of firms under pressure from the NGOs, *SURED 2018 - Monte Verità Conference on Sustainable Resource Use and Economic Dynamics*. https://ethz.ch/content/dam/ethz/special-interest/mtec/cer-eth/resource-econ-dam/documents/research/sured/sured-2018/93-SURED2018_Brecard_Chiroleu-Assouline.pdf.

García-Gallego, A. and Georgantzís, N. (2009). Market effects of changes in consumers' social responsibility, *Journal of Economics & Management Strategy* **18**(1): 235–262.

García-Gallego, A. and Georgantzís, N. (2011). Good and bad increases in ecological awareness: Environmental differentiation revisited, *Strategic Behavior and the Environment* **1**(1): 71–88.

Li, Y. and van 't Veld, K. (2015). Green, greener, greenest: Eco-label gradation and competition, *Journal of Environmental Economics and Management* **72**: 164–176.

Manasakis, C., Mitrokostas, E. and Petrakis, E. (2013). Certification of corporate social responsibility activities in oligopolistic markets, *Canadian Journal of Economics/Revue canadienne d'économique* **46**(1): 282–309.

Miller, D. T. and Monin, B. (2016). Moral opportunities versus moral tests, *in* J. P. Forgas, L. Jussim and P. A. Van Lange (eds), *The Social Psychology of Morality*, first edn, Routledge, pp. 40–55.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality, *Journal of Economic Theory* **18**(2): 301–317.

van 't Veld, K. (2020). Eco-labels: Modeling the consumer side, *Annual Review of Resource Economics* **12**(1).

van 't Veld, K. and Kotchen, M. J. (2011). Green clubs, *Journal of Environmental Economics and Management* **62**(3): 309–322.